University of Alberta

ON CONTENT AND TRUTH-CONDITIONS

by

Tal Miller Aviran

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

Department of Philosophy

Edmonton, Alberta
Spring 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-82080-7
Our file Notre référence
ISBN: 0-612-82080-7

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

University of Alberta

Library Release Form

Name of Author: Tal Miller Aviran

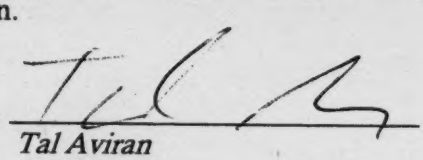
Title of Thesis: *On Content and Truth-Conditions*

Degree: *Doctor of Philosophy*

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publications and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



Tal Aviran
32 Sirkin St.
Herzelia
Israel, 46392

Submitted to the Faculty of Graduate Studies and Research 12/23/2002

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *On Content and Truth-Conditions* submitted by Tal Miller Aviran in partial fulfillment of the requirements for the degree of PhD.

Dr. F. Jeffry Pelletier

Dr. Bruce Hunter

Dr. Alexander Rueger

Dr. Robert A. Wilson

Dr. Michael R. W. Dawson

Dr. Joseph Almog

Supervisor writes the date that the thesis is approved by committee here 12/20/02

Abstract

The purpose of this dissertation is to come-up with a theory of mental content as a causally efficacious property. It is claimed that such a notion of mental content needs to satisfy seven desiderata: A) mental content has to be a causal-nomic property, that is, one which is subsumed by causal-laws. B) mental content has to determine causal processes which are isomorphic to its semantic liaisons. C) mental content has to be an atomistic or molecularistic property: it cannot be a holistic property. D) mental content has to be individuated at an empirically adequate level of grain. That is, the theory has to account for the empirical possibility of informational differences between coextensive terms (Frege's condition), and for the empirical possibility of theoretical substitutions of terms that are neither coextensive nor co-referential (Loar's condition). E) mental content has to be a representational property that is able to encode information about objects and states of affairs in the world. F) mental content has to be realized by a supervenience base which could endow the content-bearing states of physical Twin- earthlings with distinct causal powers. G) mental content must allow for the possibility of error or misrepresentation.

I argue that a theory of Modal Response Dependence Informational Content is the theory which can satisfy desiderata A-G. As informational content is constituted by nomic relations to the distinct instantiations of properties in the world, it is nomic, atomistic, representational, and can satisfy the supervenience thesis by reference to relationally individuated brain-states. It also satisfies the isomorphism condition as content is identified with causal relations to objects and events in the world.

Construed in terms of informational relations, the theory of content becomes a species of response-dependence theories of content, with content

relativized to selective responses of the organism to objects and events in its environment. However, if limited to actual responses, those informational relations turn out to be either too coarse or too fine to ground mental content. It also threatens the account on error. It is therefore suggested that content be determined by reference to *possible* selective responses (or discriminations) of the organism, with the modality ranging over the set of cognitively possible worlds. I argue that such a theory can then satisfy the remaining desiderata on grain and error.

Acknowledgments

I began the actual writing of this thesis in October 2000 in Kibbutz Ga'aton in Israel, and finished it in December 2002 in Edmonton Alberta. Within this time period I have written quite a number of drafts of the chapters which found their way to the final version of the thesis, and perhaps even more drafts of chapters which did not. All these drafts were read and commented on with meticulous diligence and for the most part with a lot of patience by my supervisor Professor Jeff F. Pelletier. Most of all I want to thank Jeff for the intellectual support and freedom of creation he has given me throughout the writing process and in fact, since my first arrival to the university of Alberta. I don't think I could have wished for a better supervisor than him.

In the second year of writing the thesis professor Robert A. Wilson entered the fray when he replaced a departing member on the supervisory committee. Very soon after, Rob turned out to be a major driving force in pushing the thesis a few notches up, and as far as I'm concerned, became in effect a second supervisor to the thesis. It is impossible for me to describe here the amount of work that Rob put into this thesis, with most of which, by the way, I believe he strongly disagrees. Although there were times in which, I must say, I was exasperated by his perfectionism, I am reluctant to think how my thesis would have developed without the contribution of Rob's knowledge and brilliance.

Aside from Professors Pelletier and Wilson, I also received a lot of support from my wife Audrey. First, Audrey had to hear any new idea that came to my mind before it went into a draft form, and I must say that some of these discussions helped me understand better my own thoughts. Second, Audrey also helped with editing the bibliography of the thesis and other editing tasks, and in

drawing some of the graphical depictions, although unfortunately not all of them found their way to the final version. But most of all, Audrey gave me the confidence to finish the thesis at times when I was ready to drop everything and return to my old job as a taxi driver in New-York City. I might be able to do it now, though.

Before arriving to the university of Alberta I was a Ph.D. student at the Hebrew university of Jerusalem, where I started work on a thesis in ancient philosophy titled: "Plato's Parmenides". My three and a half years at the Hebrew university were definitive to my philosophical development, as I had the opportunity to work with some of the greatest philosophers of our time, Professor S. Scolnicov and Professor Y. Yovel. From professor Scolnicov, who was my ancient philosophy supervisor and teacher, I learned most of what I know today on Plato, which, though, not an "analytic" philosopher, is still for me the greatest philosopher ever. From Professor Yovel I learned that philosophy is about wisdom as much as it is about acumen. Professors Scolnicov and Yovel helped me turn from an enthusiastic B.A. graduate to a nascent scholar. Although my thesis is in a field in philosophy which is somewhat far-off from their interests, I believe that their influence on me can be felt at every turn of page.

For a long time the major figure in my philosophical as well as personal development was my beloved friend Tal Kohavi. The influence that her style of thinking and way of being had on me and on my doing philosophy goes beyond what I can convey here. Unfortunately, Tal chose not to be there when I wrote my thesis. Her wisdom and support were profoundly missed.

Last but not least, I want to thank my parents who had to wait patiently all these years to see their son finish his Ph.D., and in philosophy of all things. I hate to think about the amount of embarrassment that my parents must felt

whenever they had to tell their friends that their son, already in his forties, is still a student. Well, no more.

TABLE OF CONTENTS

Introduction 1

Chapter One

Reasons, Causes, and Intentional Laws 7

1. Intentional Realism 7

2. The Representational Theory of Mind (RTM) 8

3. The Nomic Character of PEPs 10

4. The Plausibility of Psychological Laws 12

5. The Objection From Normativity: Reasons Cannot Be Causes 16

6. The Problem of Explanatory Force 18

7. Problems For Davidson's Model 21

8. Two Theories of Explanation 22

9. The Theory of Modal-Disjunctive Properties (I) 25

Chapter Two

The Isomorphism Thesis (IT) 34

1. Isomorphism 35

2. Conceptual Role Theory (CRT) 37

3. On Functionalism and Intentional Realism 44

4. The Problem of Fusion 46

5. The Argument from Holism, or Meaning Incomparability 47

6. Ideal Verificationism (or 'Idealism') 53

7. The Argument from the Underdetermination of Content 57

A. The Problem of the Generality of Laws 58

B. The Problem of Equivalent Propositional Attitudes 59

8. Content Is Not Inferential Role 60

Chapter Three

The Theory of Cognitive Values 63

1. Some Presemantic Considerations 63

2. Frege's Theory of Meaning 67

3. Frege's Solution to Frege's Puzzle 70

4. Russell's Solution to Frege's Puzzle 74

Chapter Four

Propositionalism, the Dyadic Theory of the Attitudes, and a bit more on Frege's Puzzle 83

1. Intentionality and Representation 85

2. On the Logic and Semantics of 'That'-Clauses 89

3. Frege and Russell's Theories of the Propositional Attitudes 93

4. Bealer's Theory of Propositions 98

5. On Conception 1 and Conception 2 Entities 100

6. Naming trees, Causal Theories and Non-Platonic modes of

Presentation 104

7. On Propositions and Logical Closure 111

8. Mates' Criterion and the Method of 'Psychoanalytic Ontology' 114

Chapter Five

Individualism 120

1. A Non-Intentional Core 120

2. Two Individualistic Theses 124

3. Twin-Cases 130

4. Narrow Content 136

5. On Narrow Content and Causal Powers 144

6. On Relational Taxonomies and Causal Powers 150

7. On Causal Powers and Supervenience 156

Chapter Six

Informational Semantic Theory 161

1. Preview of the Informational Concept of Content 161

2. The Purely Informational Semantic Theory 172

3. The Disjunction Problem 176

4. Fodor's Asymmetric Dependency Solution 179

5. Property Theory and Implementing Mechanisms 181

6. The Theory of Informational Concepts 191

7. On Narrow Content and Disjunctive Properties 196

8. Summary: The Definition of Mental Content 199

Chapter Seven

The Modal Response Dependence Informational Theory 202

1. Précis of Previous Chapters 202

2. The Informational Theory of Content 210

3. The Theory of Modal Disjunctive Properties 216

Bibliography 222

List of Figures

Figure 1.1: The implementation of "higher-level" processes by "lower-level" mechanisms 27

Figure 6.1: The implementation of intentional laws by lower-level mechanisms 182

Figure 6.2: The implementation of the concept HORSE. 185

Figure 6.3: The dependence of 'horse' on HL rather than on either horses or cows on a dark night 197

Figure 6.4: 'horse' means P(INF) 198

Figure 7.1 Modal disjunctive properties 218

List of Acronyms (in order of appearance)

PEPs	Psychological Explanations and Predictions	8
RTM	Representational Theory of Mind	8
FP	Folk Psychology	8
PAM	Principle of the Anomalism of the Mental	13
IT	The Isomorphism Thesis	15
CRT	Conceptual Role Semantic Theory	34
RTT	Representational Theory of Thought	47
LOT	The Language of Thought Hypothesis	47
PLP	Principle of the Logical Propriety of Expressions	77
MS	Methodological Solipsism	124
WL	Water-lookingness	170
IC	The Informational Condition	174
RC	Robustness Condition	178
AD	Asymmetric Dependence Condition	179
HL	Horse-lookingness	185
AC	The Actualist Clause	187
P(INF)	The Disjunction of all the proximal stimuli of horse	197
HD(INF)	The disjunction of all the Horse Detection properties	198
CD(INF)	The disjunction of all the Cow Detection properties	198
MRDSC	Modal Response-Dependent Supplementary Conditional	199
MC	The Informational Definition for Mental Content	212

Introduction

In this dissertation I attempt to give an account of mental content in terms of a finely-grained notion of truth-conditions, or what I take to be the same, information content. I argue that this notion of content is required in order to support what is arguably the central tenet of intentional realism: the idea that mental content is a causal property of intentional mental states which determines their causal role. It should be noted that I make no attempt to argue *for* intentional realism but merely to answer the quasi Kantian question of *how is it possible* for a mental state to have a causal-role *as* an intentional state, i.e., in virtue of its content. Indeed, to supply an answer to the Kantian question is what I see as the thesis' objective. In order to determine when such an answer has been given, I have formulated a list of seven desiderata which, I claim, must be satisfied to achieve the task. These desiderata should be seen as a set of conditions each necessary and jointly sufficient to answer the Kantian question. I will first formulate the desiderata in a somewhat rough form, and then explain them in more detail. This will also give me the opportunity to introduce the chapters which follow.

Desiderata on mental content:

A) The Nomic Condition: mental content has to be a causal-nomic property, that is one subsumed by causal-laws.
B) The Isomorphic Condition: mental content has to determine causal processes which are isomorphic to its semantic liaisons.
C)The Atomistic Condition: mental content has to be an atomistic or at most a molecularistic property. It *cannot* be a holistic property.
D) The "Grain" Condition: mental content has to be at an empirically adequate level of grain. This condition can be subdivided into two sub-conditions:

> D1) Frege's Condition: A theory of Mental Content has to solve Frege's puzzle about identity and the problem of opacity more generally (involving co-referential proper names and singular terms, coextensive general terms, and equivalent terms and propositions). That is, the theory needs to account for the possibility of differences in information content between coextensive terms.
>
> D2) Loar's Condition: A theory of Mental Content needs to explain why some relational individuations of mental states are too finely-grained, i.e., why there is no empirical problem of theoretical substitutions of terms which are neither coextensive nor co-referential.

E) The Truth Condition: mental content has to be a representational property. It has to be able to give information (i.e., truth) about objects and states of affairs in the world.

F) The Supervenience Condition: mental content has to be 'realized' by a supervenience base which allows the content-bearing states of twin earthlings to have distinct causal powers.
G) The Error Condition: mental content must allow for the possibility of error or misrepresentation.

Desideratum A is the requirement that for content to be a causally efficacious property, there must be intentional causal laws that subsume it as such. It is derived from the Humean conception of causation as subsumption under laws. In chapter One we shall see that one strong motivation for the belief that intentional mental content satisfies A is the seeming causal-nomic structure of psychological explanations and predictions as practiced by the Folk (in 'Folk Psychology'), and also by Cognitive Science. However, later in chapter One I shall also consider an a priori argument according to which intentional states cannot enter into causal relations with the very same states they "rationalize". This is the argument that Reasons cannot be Causes, which expresses another Humean conception, this time that causes must be logically independent of their effects. My claim will be that Reasons *can* be Causes if there is a way to make the reason metaphysically independent of its effect in *modal* space. This is where I introduce for the first time the theory of modal-disjunctive properties: the idea that the instantiation conditions on properties, pertaining to which property was actually instantiated, include actual as well as counterfactual conditions.

Desideratum B, concerning the isomorphic relation between mental processes and the semantic liaisons between their intentional descriptions, is introduced in chapter One under the name 'the isomorphism thesis' (IT), but is discussed more extensively in chapter Two. Satisfying IT requires a solution to the reasons-causes chestnut, but also to the more general problem of isomorphism as such. In chapter Two I argue that while a prevalent notion, isomorphism lacks the required determinacy to constrain the causal relations between intentional mental states by reference to their semantic liaisons. Roughly put, the problem of isomorphism is that with a bit of fiddling, anything can be made isomorphic to just anything else. An attempt is then made to substantiate IT by reference to a 'use' theory of content, such as Conceptual Role Semantics (CRT). CRT aspires to define the content of intentional states, given in terms of their inferential-roles, by reference to their causal roles. If the inferential liaisons between the attitudes are just identical to their causal liaisons, then IT would be satisfied trivially. But as the discussion shows, CRT suffers from some serious problems of its own. For one, it is holistic and thus introduces the problem of meaning incomparability. If the content of any belief depends on everything else the agent believes, desires, intends, etc., then it is very unlikely that any two agents could share a belief with the same content (or even that the same agent

would instantiate two states with the same content at different times in her cognitive history). Thus arises desideratum C that mental content be an atomistic property. But holism also has implications to desideratum A, since laws require generality whereas content holism make content relative to particular cognitive systems. Other problems with CRT that I shall discuss are that it leads to the "fusion" of contents to their bearers (i.e., no structured thoughts), and that it makes content too "coarse" from an empirical point of view. The latter is connected to the complaint that content is more "fine-grained" than inferential role: states or propositions which enter into the same inferential relations can still determine different contents, as the empirical data seems to show. Our conclusion from chapter Two will be that CRT cannot substantiate IT in a way that gives us an atomistic notion of content with enough grain to be empirically adequate. I will argue that we require a notion of content which determines inferential-role and then causal role, rather than the other way around.

The issue of grain is picked-up again in chapter Three in the context of the so called 'Frege's puzzle' and desideratum D1. Roughly put, the question that 'Frege's puzzle' raises is what makes for the distinct cognitive values of coreferential and coextensive terms. If by 'cognitive value' one understands the information content of such terms (an interpretation I will argue for in the text), then the conclusion leads in the direction that such terms must have distinct *truth-conditions*. This is one motivation for desideratum E that content has to be truth-conditional, or representational content. (The other motivation for E is that we tend to think of the expediency of content in terms of its informational value to the organism vis-a-vis the environment in which it survives and proliferates). But unlike 'extensional' theories of truth-conditions where substitution of co-referential terms should make no difference to the truth-conditions of the sentence, an informational solution to Frege's puzzle requires an 'intensional' conception of truth-conditions. That is, I will claim that satisfying desideratum D1 requires us to assign distinct truth-conditions to (e.g.,) 'water is wet' and to 'H2O is wet', to 'the Morning Star is far' than to 'the Evening Star is far' and even to 'Cicero is smart' and to 'Tully is smart', consonant with Fregean semantics. In fact, my claim will be that if Fregean semantics assigns the same content to synonymous (i.e., equivalent) terms, we need pursue an even finer scheme of content individuation than Frege allows, by reference to what I call Mates' criterion (discussed in chapter Four section 8). That is, I shall argue that Mates' criterion supplies a notion of content which is *just* empirically adequate, in that it reflects (but does not constitute) the functional role of concepts in human psychology.

The idea that content is truth-conditions makes the determination conditions on content 'external' to the agent. This raises two problems, one concerning the possibility of fine-graining content too much, the other concerning the 'internal' realization of content as a causally efficacious property.

These two issues bear respectively on desiderata D2 and F, and relate to the so called 'Twin cases' introduced by Putnam and Burge. Both Putnam and Burge claimed that individuals who are physically identical in all their intrinsic physical properties can still entertain mental states with distinct contents, given differences in either their physical or social environments. Putnam even claimed that while the 'what's in the head' of the Twins is the same, the content of their mental states could still differ. To show this, Putnam has presented a case where two physically identical Twins are situated in distinct physical environments. In particular, the micro-structure of the stuff they both think about and describe as 'liquid, potable, transparent' etc., and for which they both use the vocable 'water', is H20 in the one case, and XYZ in the other. Because of these differences, Putnam claimed that the contents of their relevant 'water'-thoughts differ respectively in that it is H20 in the first case, and XYZ in the second case.

Putnam's theory supplies an extensional conception of content if one takes 'water' and 'H20' to be coextensive terms in English, and 'water' and 'XYZ' as coextensive terms in Twin-English. It is also a causal-historical theory of content as it takes events in the history of the individual's acquisition of her concepts (or meanings) as determinative of the contents of her thoughts (words). Had we changed the history of the Twin's conceptual acquisitions, we might have changed the contents of their mental states, and that without any changes in their physical constitution or in their relevant conceptions (the *what* they think concerning water/twater). But the problem with this theory is that such differences in content do not always make a difference to our psychological theorizing vis-a-vis the Twins. As Loar made the point, it seems that we can substitute in a psychologically adequate theory any occurrence of a truth-conditional (or "wide") content like *H2O* for that of *XYZ* and vice versa, without affecting the empirical adequacy of the theory. From which we should conclude that individuating 'water'-thoughts by reference to their extension - the objects or stuff the thought is about - might carve contents too finely. Desideratum D2 requires that our theory of mental content respects the fact that at times, systematic substitutions of *non-coextensive* terms do not affect psychological adequacy.

Loar's condition also bears on the 'local' causal efficacy of truth-conditional content. If truth-conditional content is externalist in that it determines content by reference to conditions outside the individual, then the question arises about the causal efficacy of that property with respect to other 'local' states of the organism: other mental states as well as behavioral and perceptual states. Put somewhat differently, the question is how the truth-conditions of a state, a semantic property of a state which has to do with its temporally and spatially unbounded "aboutness", can make a difference to its causal powers in the here and now. As we shall see in chapter Five, questions about causal efficacy of higher-level properties are commonly answered by

reference to some supervenience relation to physical properties of the organism. However, this neo-positivistic conception, as I call it, implies a notion of content which is dependent on states of the organism rather than on states of its environment. My take on this will be that there is an alternative notion of a supervenience relation which is consistent with truth-conditional semantics for mental content, and therefore with the intuitions derived from the Putnam and Burge thought-experiments. This is a notion of a supervenience relation in which the supervenience base is a set of brain states themselves widely individuated by reference to their causal-informational relations to objects and events in the organism's environment. In this way, local causal efficacy is ensured, while physically identical organisms can instantiate states with distinct causally efficacious contents. Provided, that is, that those states are subsumed by distinct informational laws to distinct environmental instantiations. Consequently, the truth-conditional theory of content becomes the theory of informational content.

The attempt to construe content in terms of information requires more than covariance relations. It has to account for the possibility of error as well. While a signal carries information about anything which covaries with it, the same cannot be true for semantic content, as shown by the disjunction problem. To get a natural occurrence of error, a content bearing state (such as 'x') would have to be caused not only by the content property 'X', but also by non-X's (i.e., Y's). However, if 'x' covaries alternatively with instances of X and Y's, the causal theory determines that its content would be (*XvY*) rather than *X*. In which case we do not get error after all, only disjunctive contents. To solve the disjunction problem, some asymmetry has to be found between veridical occurrences of a symbol and its wild occurrences. This is the topic of chapter Six. I begin the discussion of the disjunction problem by focusing on Fodor's asymmetric dependency solution. Fodor claimed that the required asymmetry can be achieved by making the laws that subsume the wild occurrences dependent on those which subsume the veridical occurrences while the latter are independent of the former. More specifically, it requires that there would be possible worlds in which only X's cause 'x's, and worlds in which both X's and Y's cause X's. I claim, though, that the possibility that not only X's cause 'x's but also Y's strongly suggests that this is due to something that X's and Y's share, something to which the *mechanisms* for the detection of X's are sensitive. But then, any change in the mechanism that would affect the responses of 'x's to Y's would have a similar effect on its responses to X's. This implies that a world in which Y's cannot cause 'x's will also be a world in which X's cannot cause 'x's.

My solution to the this problem is that rather than trying to establish an asymmetry between the wild occurrences of 'x' and its veridical occurrences by affecting the relation between X's and non-X's, we should affect the relation between 'x's and non-X's. This we can do, I argue, by stipulating possible worlds in which non-X's are not sufficient for instantiating that property which in other

possible worlds they share with X's, and to which 'x's respond . In effect, this just means stipulating possible worlds in which Y's are distinguished from X's, and this without necessarily affecting a substantive change in the physical constitution of the organism. The idea is to rely on methods of selective detection of X's which can be "quantified-over"; i.e., methods which form part of the channel conditions for the 'flow of information', rather than the information content itself. I call the theory that backs-up this solution 'Modal Response Dependence Informational Theory', and the condition of asymmetry it formulates, 'Modal Response Dependence Supplementary Condition' (MRDSC). I then incorporate this condition into the definition of mental content which I give at the end of chapter Six.

Thus, the theory of content that I develop in this dissertation is one of informational/truth-conditional content where content is determined by reference to a discriminative, covariance, relation between states of organisms and the instantiation of properties in the world. Because such patterns of covariance relations are specified modally rather than actually (or historically), I construe the content properties with which informational states of the organism covary as disjunctively based properties. In chapter Seven I go-on to give a short description of this theory whose full development I leave for another occasion.

Chapter One

Reasons, Causes, and Intentional Laws

1. Intentional Realism

Any theory of mental content within the framework of intentional realism needs to have the resources to account for the role of content in mental causation. As intentional realism is the view that intentional-mental states and events are causally efficacious, indeed, that they are such *in virtue* of their *content*, then a theory of *intentional causation* is called for.[1] What we seek is a theory which would purport to answer the somewhat Kantian question of 'how is intentional causation possible?'. How is it possible for a mental state to have a causal-role *as* an intentional state, that is, under its intentional description? I note here that this question should be distinguished from the quite different one of how it is possible for a mental state to have a causal role under *some* description, for example, under its physicalistic description In what follows, we shall have the opportunity to see how these two questions can receive different answers, and also why answering the second is far from adequate as far as answering the first. The considerations invoked in this chapter are mainly in the fields of the philosophy of psychology and the theory of action, but their implications are such that any theory of intentional causation needs to adopt them.

In this chapter my purpose will be to begin to build some motivation for the kind of theory of mental content that is required, so I claim, by any theory of intentional causation. This is a theory of content which draws on a fine-grained theory of properties to make for the distinction between the so-called extensional and intensional components of content. The question of intentional causation will be treated via a related issue: that of the status of Psychological Explanations

---

[1]Two remarks. 1) The notion of 'in virtue of' is meant to be used here in its Quinean interpretation, namely as "almost a 'because of' " (Quine 1974: 8-9). Thus we can say that the table is wooden in virtue of, or because of its having a certain material constitution, or that Bob became an uncle in virtue of, or because his brother became a father (see also McLaughlin, 1989: 114-5). 2) With respect to the notion of intentional causation, I want to distinguish my use of the term from that of Searle (1983). For one thing, Searle predicates that property only of perceptual states and intentions, whereas I predicate it of any intentional state which has a causal role. Second, for Searle intentional causation is not so much a species of causal relations as the condition that perceptual states and intentions be causally related to their satisfaction condition.

and Predictions (henceforth, PEPs) as practiced by the Folk. We shall encounter two opposing views: According to the first view, exemplified in Jerry Fodor's Representational Theory of Mind (RTM), PEPs are a species of nomic explanatory and predictive practices. According to the other, PEPs are normative rather than descriptive. We shall see that this latter contention gives rise to the putative conclusion that intentional states cannot be causally efficacious, and therefore, that reasons (which are intentional states), cannot be causes. At the heart of this view figures an old stricture which goes back to David Hume, according to which causes must be logically independent of their effects. This consequence could raise some serious difficulties for Fodor's theory and with it, for intentional causation generally. One attempted solution to the problem of how reasons can be causes was suggested by Donald Davidson, and I shall discuss it at length. However, we shall see that Davidson's solution rests on a questionable distinction between description *kinds*, and perhaps even on the infamous analytic/synthetic distinction itself. In the closing part of this chapter I will present my own solution in terms of a psychological theory of *intentional instantiation,* which draws on a unique metaphysical theory of 'access' properties.

2. The Representational Theory of Mind (RTM)

RTM, as Fodor admits in quite a few places (1978, 1994, 1998), is a theory that still lacks a canonical formulation. It is more a budget of theses for which it is not completely clear how, even at all, they hang together. Some reasons for that might be traced to the origin of the theory in Folk Psychology (henceforth, FP), that is, in the set of principles and concepts used by the Folk to explain and predict the behavior and mental states of other people and their own. Indeed, Fodor sees RTM as a scientific vindication of FP, drawing both on its ontology and on its platitudes (i.e., the commonsensical generalizations). Examples of such platitudes might be that people usually intend what they say, that they usually try to get what they want, and even that men try to avoid marrying their mother (under the relevant description). RTM is one attempt to give a scientific format to such principles, accepting thereby that the underlying structure of FP is theoretical. One implication of this is that the Folk generalizations are defined over unobservables.[2] Here we come to the ontology of the theory: the contention

[2]Of course, there are other options, as that of reducing (or at least replacing) talk about intentional states to talk about observable states or dispositions to produce such states (see e.g., Ryle 1949; Skinner, 1953, Davidson 1990). Opponents of such attempts usually point to the failure to give a satisfactory complete and non-circular reduction of mental states in term of

that the folk generalizations are true of states which are 1) semantically valuable[3], that is, they have a proposition as a (direct or indirect) object[4] and 2) they have causal powers. These are the propositional attitudes, common examples of which are beliefs and desires (hence the locution 'belief/desire psychology') (Fodor 1987:9, Davidson, 1963/1980 Essay 1).[5]

By advancing RTM, Fodor not only places himself in the intentional realist camp, that is, amongst those philosophers who accept the ontology and generalizations of FP, but he wants to go beyond it, into scientific intentional realism. In the next section we shall see what this entails, but here I should note that although RTM was presented as an *empirical* theory, it is conceded by Fodor that there seem to be some conceptual issues that threaten it. The problems are serious, it should be noted, since they touch upon the most sensitive part of the theory: the part where the psychological meets the semantical. In fact, it is these conceptual problems that will occupy us in what follows.

At the heart of RTM are three theses[6]:

A) The thesis about the Nomological-Intentional character of Psychological Explanations and Predictions.

B) The thesis about the Representational and Computational character of mental states and processes.

C) The thesis about the informational nature of mental content.

---

observables, and the need, in any case, for unobservable intermediaries to supply satisfactory explanations and predictions of behavior (see Chomsky 1959; Fodor 1968a; 1968b).

[3]I prefer the locution 'valuable' to '*e*valuable' since the latter implies the activity of an assignment of values, something which I think is irrelevant to the semanticity of mental states or events.

[4] I leave it open here whether these propositions are Fregean, Russellian, or whatever. I shall get back to this in chapter Three.

[5]Perhaps a stronger motivation for the locution 'belief-desire' psychology came from those who hoped to reduce *all* intentional states to these two kind of states. This project, however, seems to be facing some serious difficulties.

[6]A methodological point: Since to each and every thesis of RTM as presented here there is as much dissent as there is assent even *amongst* proponents of RTM, I will stick in my discussion to Fodor's version of it. Objections to various aspects of Fodor's theory will be considered as they become relevant to the discussion.

Thesis A is the claim that psychological explanations and predictions involve the derivation of mental and behavioral events from causal-laws under their intentional descriptions, that is, by reference to their propositional content. Thesis B is the claim that mental states are relations to mental representations as syntactical structures with semantic content, and that mental processes involve manipulation of these symbols according to a predetermined set of rules. Thesis C concerns specifically the content of mental states. The claim is that the reductive conditions for content (that is, its naturalistic base) are to be given by reference to informational relations; presumably, between some neural states of the organism and objects and states of affairs in the world.

In the rest of this chapter, dedicated as it is to the theory of action, I will analyze in the main the relation between thesis A and the contention that psychological explanations and predictions are normative rather than descriptive. This will also give us the opportunity to review some of the arguments in favor of the nomic nature of mental-intentional causation. Theses B and C will be discussed in the following chapters.

3. The Nomic Character of PEPs

In Fodor (1998:7), we find the following pronouncement of the nomic character of PEPs:

> *Psychological explanation is typically nomic and is intentional through and through.* The laws that psychological explanations invoke typically express causal relations among *mental states that are specified under intentional description*; viz. among mental states that are picked out by reference to their content. Laws about causal relations among beliefs, desires, and actions are the paradigms.[7]

Let's see what Fodor is claiming here. As we saw above, FP explanations and predictions attempt to account for past or future behavioral and mental occurrences: 'Why did Brutus stab Caesar?', 'Why did Hermia think that Demetrius killed Lysander?', and perhaps more mundanely, 'what would Mary do when the stop-light turns green'? Fodor thinks that what is right about the Folk PEPs is that they implicitly rely on the structure of scientific causal explanations.[8] One aspect of this, which is mentioned above, was that they range

[7]In this work italics are in the original unless stated otherwise.

[8]Surely, not all realists about the mind need believe that mental states are causally efficacious states. For a dissenting view, see Dennett 1987, and Bennett

over theoretical entities, the attitudes. Another aspect of it lies in the assumption that the contents of these states play a nomic regulative role in the way these entities interact - among themselves, and with perceptual input and behavioral output. Thesis A is just the application of these assumptions to PEPs. More specifically, it comprises the following two clauses:

1) Psychological explanations are nomic: they follow the Hempelian covering-law model of scientific explanations.
2) Psychological explanations are intentional. They subsume events under their intentional descriptions.

The first clause is an application of the more general model of scientific explanations to psychology, as introduced by Hempel and Oppenheim (1948). At the heart of this model is the claim that scientific explanations and predictions consist in the derivations of event descriptions, the explanandum, from law sentences (i.e., sentences expressing laws) plus antecedent conditions, the explanans. (Since Hempel and Oppenheim thought that the right sort of derivation is deductive, this model is also called the deductive nomological model, or D-N for short). Thus, if L is the law: 'C→E',[9] then given an occurrence of C, an occurrence of E could then be explained by showing that E is derivable from L and C. I should note that although a similar procedure is good for predicting E, there is an important difference in that a future or merely possible event can be "overdetermined" in a way in which an occurring event cannot. I will return to this point later.

Now, if we combine clauses 1 and 2 we have the consequence that PEPs involve the entailment of psychological explananda from psychological explanantia, under intentional descriptions. Let's take a closer look at that claim by comparing it to an example from physics (after Churchland, 1979).

P) (x)(f)(m)[((x has a mass of m) & (x suffers a net force of f))→ (x accelerates at f/m)].

Here the expressions 'has a mass of m' or 'accelerates at f/m' are predicate forming expressions in that a substitution of a numerical value over which the quantifiers range will result in a determinate predicate (e.g., has a mass of 2 kg). In this way, an explanandum could be entailed by making the appropriate substitutions. Now, the idea is that the same goes for PEPs, only that here the

---

1990.

[9]I adopt Fodor's arrow sign for the subjunctive conditional that expresses laws. I find Churchland's use of the material conditional (the horseshoe) to be a mistake.

variables are taken to range over propositions, instead of numbers:[10]

PL: (x)(p)(q)[((x desires that p) & (x believes that (not-p unless q) & (x is able to bring it about that q))> [Barring conflicting desires or preferred strategies] (x will behave in a q manner).

Let's try this on an example:
Q. Why did Mary take the 5 O'clock train to Boston? Exp.: because she desired to get to Boston at 12 p.m., and believed that if she took the 5 O'clock, she would (probably) get there at that time, and (not less importantly) because the following conditional is an instantiation of a confirmed law (PL):

> 'If Mary desires to get to Boston at 12 p.m., and Mary believes that unless she takes the 5 O'clock train she would not get there at that time, then, barring some conflicting believes and desires (such that the 5 O'clock is always late etc.), Mary would take the 5 O'clock train'.

4. The Plausibility of Psychological Laws

According to intentional realism, the success of the Folk PEPs is to be explained by reference to an underlying, scientific-like, system of laws, from which mentalistic singular causal statements can be derived. This seems to have the implication that mental states can be considered as a natural class or kind, with perhaps discoverable scientific essences. After all, if there are *some* laws relating the mental to the physical, as the laws governing perception and behavior, then it would only make sense that there are other such laws which make for the physical realization of that capacity.[11]

---

[10]It is assumed here that such constructions are to be taken in their objectual interpretation, in which case the terms for propositions (and for numbers) are to be taken as singular terms. Of course, this point underlies the relational theory which will be discussed in chapter Four.

[11] There might be those who would like to question the whole notion of the connection between property instantiation and laws. It is hard, though, to clearly understand what these philosophers have in mind in that objection. Do they object merely to the idea that an instantiation of a property by a substance at a time implies the existence of some general regularity, that is, insisting thereby on a singularist conception of instantiation, or do they rather want to make the stronger claim that there is nothing - neither regularity in some frame of time and space, nor something internal to the substance - which underlies that

According to Donald Davidson, though, the principle just implied, that of the causal interaction of the mental with the physical, is in tension with another principle he thinks we are obliged to accept - the Principle of the Anomalism of the Mental (PAM). According to PAM, "...there are no deterministic laws on the basis of which mental events can be predicted and explained" (Davidson 1970, in Davidson 1980:208). This principle implies that there are no psycho-psycho laws, in addition to there being no psycho-physical laws: laws that reveal the scientific essences of mental properties. What makes Davidson think that PAM is true? In effect, Davidson advances two arguments for this principle. One argument, which is somewhat Quinean in nature, is based on the claim that mental ascriptions are normative and hence lead to a certain indeterminacy for mentalistic predications. This argument, which is the stronger of the two in my view, will be discussed later in the chapter. The other argument Davidson has for PAM has to do with the "hedged" character of psychological laws. It is basically the complaint that "the generalizations that embody such practical wisdom [i.e., FP] are assumed to be only roughly true, or they are explicitly stated in probabilistic terms, or they are insulated from counterexample in lieu of generous escape clauses"(1970/1980:219). That is, the claim is that psychological laws are 1) not precise enough to count as *causal* laws and 2) are not refutable.[12] How strong is that objection?

It has been argued, most vehemently by Fodor (1974, 1989, 1991)[13], that the idea that only strict (i.e., exceptionless) laws should count as genuinely scientific, would exclude not only psychology from the realm of the scientifically kosher, but also many other candidates for the title. Thus, Fodor has long claimed that psychological laws are no different than other "special" science

---

instantiation. If the former, than the objection is not really to laws but to the idea that such regularities constitute laws. If the latter, then it would seem that there is nothing left to explain instantiation. I myself am willing to follow Dretske (1977), Tooley (1977) and Heathcote and Armstrong (1991) in seeing causal instantiation, and in fact, any instantiation, as a basic relation, though contingent, between properties (universals).

[12]It should be stressed that Davidson does not object to there being psychological laws as such, only to the claim that these laws can be causal laws (cf. Davidson 1993; McLaughlin 1989, 1993). The bearing of this issue on the causal efficacy of the intentional will be discussed later.

[13]Cf. Anthony 1994. Anthony also argues, relying on an example by Coffa (1973), that even some physical laws have *ceteris paribus* clauses. On a dissenting view, one which is even more radical than that of Davidson in that it repudiates the existence of *ceteris paribus* laws altogether, see Schiffer 1991.

laws, such as those of Geology, Biology, etc., in that the generalizations of the latter are no less inundated with *ceteris paribus* clauses than psychological generalizations. In this, Fodor has taken a weakened view of Hempel's model in its application to the "special" sciences, qualifying its strict application only to ideal conditions of the system. What lies behind this qualification is the claim that all hedged laws, that is, all the laws of the "special" sciences (and in fact also many of the laws of physics), need mechanisms for their implementation, and mechanisms have conditions of proper or ideal operations. The point is, however, that these conditions can be specified only at a lower level than that of the laws, as the following example, one of Fodor's favorites, shows: *'ceteris paribus*, Meandering rivers erode their outside banks', *unless*, that is, their banks are made from concrete, iron, etc. But 'being made from concrete' is not a term in geology and has rather to do with the particular realization of the events subsumed under the law. As we shall see, RTM's (or more specifically, CRTT) uniqueness in the intentional realist camp is in its claim that the lower-level mechanisms which implement the higher-level processes are computational.

The above considerations seem to have established that if psychological generalizations are ruled-out because they are hedged, then so must be the case with the laws of all the other "special sciences". Thus, if we accept hedged laws in Geology and Chemistry then Fodor's point is that there is no reason why we should not be allowed to do the same in psychology. The underlying assumption is that there really is no difference in kind between psychological generalizations and the laws of the special sciences.

But we saw that there is one aspect of psychological generalizations which makes them stand-out from the rest, in that they not only capitalize on certain mental regularities but rely also on the logical relations between the objects of the attitudes. Thus, it was claimed that there might be even an incompatibility between the conception that psychological explanations are intentional and the contention that FP is an empirical theory; that is, that its generalizations depend on their instances for their truth. For the question arises as to how could this be, given that it seems that what makes PEPs true are just the *logical* relations between the propositional objects of the attitudes.

We can look at this point from a different perspective. To claim, as Fodor does, that the laws which subsume mental events are intentional is to go beyond the triviality that they subsume their instances under their nomic descriptions. The geological law that meandering rivers erode their banks subsumes its instances under descriptions which bear no logical relations to each other. Here the only factor relevant for this law being true is that it be confirmed by its instances and support counterfactuals. On the other hand, it seems that what makes the likes of (the psychological law) PL true are rather the relations between the propositional *objects* of the states, as expressed by the law sentences. And this gives the impression that the necessity accrued here to the conditional

is logical, rather than nomological. At the very least, it seems that we have *two* modal relations operative in cognitive psychology, and that something needs to be said about the relations between them. Fodor, indeed, sees this phenomenon as a virtue: "the parallelism between causal powers and contents engenders what is, surely, one of the most striking facts about the cognitive mind as commonsense belief/desire psychology conceives it: the frequent similarity between trains of thought and *arguments*" (1987:13). And in another classic passage, he continues:

> The basic idea is that, given the two networks - the causal and the inferential - we can establish partial isomorphism between them. Under such an isomorphism, *the causal role* of a propositional attitude mirrors the semantic role of the proposition that is its object. So, for example, there is the proposition that John left and Mary wept; and it is partially constitutive of this proposition that it has the following semantic relations: it entails the proposition that John left; it entails the proposition that Mary wept; it is entailed by the pair of propositions {John left, Mary wept}; it entails the proposition that somebody did something; it entails the proposition that John did something; it entails the proposition that either it's raining or John left and Mary wept...and so forth. Likewise there are, among the potential episodes in an organism's mental life, states which we may wish to construe as (S1) having the belief that John left and Mary wept; (S2) having the belief that John left; (S3) having the belief that Mary wept; (S4) Having the belief that somebody did something...The crucial point is that it constrains the assignment of propositional content to these mental states that the latter exhibit the appropriate pattern of causal relations. In particular, it must be true (if only under idealization) that being in S1 tends to cause the organism to be in S2 and S3; that being in S1 tends to cause the organism to be in s4 ...and so forth (1987:78-9)

I would claim that the thesis expressed in this paragraph - which I call the Isomorphism Thesis (or IT) - is probably the hallmark of RTM, and one which presents one of the most challenging problem for intentional Realism (perhaps only second to the problem of content itself). Put in a short form, IT is the contention that, either directly or indirectly, the causal relations between the propositional attitudes are *a priorily constrained* by the logical relations between their propositional objects (*ceteris paribus*).[14] If true, IT could account for the

[14]The *ceteris paribus* is here to remind us that this constraint is limited since logic can get quite complicated. In fact, that is probably where psychologism (the reduction of logic to psychology) went wrong (see also

semantic nature of mental processes; that is, of how mental processes can be said to preserve semantic properties - as that of going from true thoughts to true thoughts. But to go back to our question above, is such parallelism between the semantic and the causal even possible? That is, is it possible for the mental to be both nomic *and* logical? We shall now see some arguments to the effect that Fodor cannot have it both ways.

5. The Objection From Normativity: Reasons Cannot Be Causes

Around the middle of the last century, an a priori argument against the possibility of naturalizing the mind became quite popular. Some philosophers, e.g., Ryle (1949) Melden (1961) and Anscombe (1976), put forward the claim that the nature of psychological explanations and predictions *cannot*, by threat of conceptual confusion, be causal-nomological. The reason given was that PEPs belong to a different category than that of the laws of nature, the category of norms.[15] Since it follows from this argument that mental states cannot be causally efficacious, these philosophers are sometimes referred to in the literature as the anti-causalists (see Antony, 1989).

According to the anti-causalists then, psychological explanations and predictions do not purport to describe the actual mental causes of people's behavior, but rather to give a description that would entail the action from the agent's other beliefs and desires (and the same goes for entailing other mental states). Due to their insistence on the essential role of norms of rationality in the construction of PEPs (such as coherence, consistency, closure under entailment, etc.), this kind of position has also acquired the name Normativism (see Fodor and LePore 1991:142). To be sure, different normativists have advanced different norms of rationality, but the consequences were very much the same: mental ascriptions under PEPs should answer to norms rather than facts: they do not correspond to what mental states the agent *actually* had that had caused her to act in such and such a way, but to what mental states she *should* have had, or what action she *should* exhibit, given her presumed mental economy.

---

nonmonotonic reasoning theories).

[15] A similar objection, on methodological grounds, was raised by Dray (1957, 1963), but it is one which I will not discuss here. A more recent line of argument according to which the 'space of reasons' should be seen as *sui generis* compared to the 'space of causes' was given in McDowell 1994. McDowell sees himself as following Sellars 1956, and Davidson 1970, 1984. I shall discuss Davidson's position below. Another example of a prominent contemporary normativist is Dennett (1987, mainly chapter 8).

Someone who has accepted the basic insight underlying normativism about mental ascriptions but without its anti-causal conclusions is Donald Davidson. According to Davidson, when we explain an action by giving the reasons the agent had for it, we *rationalize* that action (1980:3). That is, we show that the action *should* be performed by an agent given that she has (a) some sort of pro attitude toward actions of a certain kind (a desire, an urge, etc.), and b) the belief, knowledge, etc., that this action is of that kind (ibid.). To use Davidson's own example, I can explain someone's flipping the switch by giving her pro-attitude, say a desire, towards the action of the kind 'turning-on the light', and her belief that this particular action of flipping the switch is also a turning-on the light kind of action (ibid., p.4).

Here we need to pay attention to two points. One is that it is essential for rationalizations that the attitude and belief be directed towards the action under a particular description of it. As the same action can be variously described, what is at issue in rationalization is that the action performed would be described in the appropriate way for it to be made reasonable, in the light of the agent's other mental states, i.e. her reasons. That is, it is required that a certain *logical relation* obtain between the action - under that description - and its reasons. This brings us to the second point. The model of rationalization Davidson has in mind is akin to that of the practical syllogism, in which the pro-attitude and belief are used as premises of an argument from which some measure of desirability of an action is deduced. But as Davidson is careful to insist, this is a weaker notion than that the action should be *derived* from the reasons, here demanding only that it be justified by them.[16]

Now, the idea that rationalizations are those explanations which involve a logical relation between explanandum and explanans (e.g., actions and reasons) seems to be in tension with the nomic model of psychological explanations. The problem is an old one and goes back to what is sometimes called Hume's stricture. This is the constraint that only logically independent events can enter into causal relations. Given the normative model of PEPs, the anti-causalists drew from this the conclusion that reasons cannot be causes. But the same point

---

[16] The difference is not trivial. For whether a reason *justifies* an action is mostly an issue of interpretation, which brings-in Davidson's interpretative perspective on mentalistic ascriptions. Similarly, we should note that notwithstanding Davidson's emphasis on the point of view of the particular agent ("a reason rationalizes an action only if it leads us see something the agent saw, or thought he saw, in his action" (1980:3)), it is reference to *ideal* agents which is presupposed here, again, from the radical interpreter's point of view. This precludes the need to introduce *mental representations* to account for the first person's perspective.

can be also made from a non-normative perspective. If reasons stand in logical relations to actions, as seems to be entailed by RTM and the isomorphism thesis, then laws which subsume reasons and actions express logical truths, in which case they surely will not be empirical. But laws are those generalizations (involving projectible properties) which are confirmed by their instances.[17] Conclusion: there cannot be a science of the mental which subsumes states and events under intentional description, *pace* NIPEP.

6. The Problem of Explanatory Force

According to the anti-causalists' account, it is necessary and sufficient for a reason to explain an action that it rationalize it. Put differently, the idea is that an action is explained once it is interpreted in such a way that it is made to cohere with the reasons. However, it seems that this model does not live-up to what we use explanations for, which is to explain why some action *actually* took place. This was Davidson's objection. He says: "a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it" (Davidson, 1980:9). The idea which Davidson seems to be expressing here is that although rationalizing is part of what is involved in explaining an action, it cannot be sufficient for it. This seems to be especially clear in cases where more than one occurring mental state rationalizes an action, but where it still seems true to say that only one of them was the actual cause.

For example, suppose that Roy believes that if the liberals were elected to power, they would enact the new education bill, and Roy wishes to see that bill passed. So if Roy votes for the liberals, his act will be rationalized by these reasons. But in addition, Roy also strongly desires to please his father, and he believes that voting liberals would make his father happy. Certainly, it is possible that notwithstanding his liberal views, it is the latter reasons that in fact caused Roy to vote liberals.

This example and others like it show that what is missing in rationalizations and required by explanations is the explanatory force, or, the etiology - a causal account of why the action *did* take place.[18] It is because of this

---

[17] The remark in the parenthesis is intended to deal with such non-projectible properties as grue and its ilk. For a definition of a projectible property, see footnote 21

[18] The anti-causalists might object here that Davidson's argument begs the question, assuming causation where there isn't any, but this seems to fly in the face of commonsense and linguistic practice. Certainly, it is not part of the *meaning* of 'explains' that what caused the action would be indeterminate

consideration that Davidson wants to show that, in fact, rationalization is a species of causal explanation. Thus, the problem of the explanatory force became the problem of accounting for the *because* in rationalizations.

Before I get on with Davidson's solution, it should be noted that, contrary to what seems to follow from Davidson's discussion, the issue he raises is not specifically one about rationalizations. Indeed, the problem should be perceived rather as a general problem that affects the D-N model of explanation itself. The reason is that it is part and parcel of the D-N model that any inference of an event from some antecedent conditions and covering principles would count as an explanation for it. Thus, Hempel (1965:419) gave the example of explaining the expansion of a metallic rod by reference to either its being heated, or its being subjected to a longitudinal stress, since the explanandum can be derived from both these explananda. As Hempel recognizes, this is just a case of "explanatory overdetermination" (to be distinguished from "causal overdetermination"), which seems to be prevalent in scientific practice. Now to be sure, there are those who have objected that explanations as the ones just cited are incomplete since they do not tell us "*why* the rod's length increased" (Kim 1989); thus, in effect, claiming that neither of the over-determining explanations have an explanatory force, as Davidson did. But then, Davidson's problem is not specifically a problem of whether rationalization is a species of causal explanation, but the more general one of explanatory overdetermination, and of the explanatory force of the D-N model. But as we shall see, it is in the *solution* to the problem of explanatory force that Davidson's ingenuity lies.

To deal with the problem of the explanatory force, Davidson proceeds in two steps. First, he distinguishes amongst the various reasons an agent had for an action the one which actually caused it: the *Primary Reason* (which consists of the pair of a pro attitude - a desire, an urge, a trait of character, etc., - and a related belief). Second, he proceeds to show how that primary reason *can* be a cause, for which purpose he turns to Hume's definition of causation (one of them anyhow).

According to Hume "we may define a cause to be an object, followed by another, and where all objects similar to the first are followed by objects similar to the second", a definition which seemed to many to have introduced a nomic conception of causation. The problem, though, is that following Hume's definition, making rationalization a species of causal explanation would imply the need for mental laws which subsume events under mentalistic descriptions. However, due to the logical relations they enter into, it might be claimed that such descriptions are unsuited for that role. But for Davidson, that should not pose a problem, for he thinks that rationalizations do not "...necessarily indicate,

---

between its various rationalizations. For a similar point, see Anthony 1989.

by the concepts they employ, the concepts that will occur in the entailed law." (Davidson 1980:17) To drive this point home, he gives the following scenario:

> Suppose a hurricane, which is reported on page 5 of Tuesday's *Times*, causes a catastrophe, which is reported on page 13 of Wednesday's *Tribune*'. Then the event reported on page 5 of Tuesday's *Times* caused the event reported on page 13 of Wednesday's *Tribune*. Should we look for a law relating events of these *kinds?*' (Ibid.)

Now to be sure, this is a bad example, since even Davidson does not think that rationalizations are on a par with constructions such as 'the event reported on page 5 of Tuesday's *Times* caused the event reported on page 13 of Wednesday's *Tribune'*. After all, we would never attempt to use such an explanation, while explanations couched in the intentional vernacular are not only widely used but are quite successful at that. But as we shall see later, this example will help us to draw an important moral, so I suggest that we merely grant Davidson's point that not just any way an event is described is nomically kosher, and continue on.

So now, armed with this insight, Davidson pays a second visit to Hume's definition of causation:

> [Hume's definition of causation] may mean that A caused B entails some particular law involving the predicates used in the descriptions A and B, or it may mean that A caused B entails that there exists a causal law instantiated by some true descriptions of A and B. (1980:16)

We can now understand why the point with changing descriptions was important. Rationalization, according to this quantificational model of causation, is a species of causal explanation, since the events picked by the rationalizing terms could be subsumed under a causal law in lieu of their *physical* descriptions. How so? Well, this is because Davidson has another theory called the token-identity theory according to which every event, including mental events, is identical to some physical event, that is, has also a physical description. And indeed, the picture that seems to follow from Davidson's discussion of rationalization is no other than his later theory of anomalous monism (AM) (in Davidson 1970/1980) according to which mental events are identical to physical events, albeit mental *concepts* (what we might call properties) are not identical to physical *concepts* (properties). I will get to that theory later on but as regards rationalization, the idea then is that psychological explanations, governed by norms of rationality, can be also causal, and hence, accepting Hume's model of causation, nomic, given that every mental event can be subsumed under a causal law under its physical description. The point is just, and this is crucial, that the *level of description* at which mental events rationalize has to be removed from

the level of description at which they enter into nomic relations. And everything would be fine so long as we keep these levels separated that way. I summarize this idea (which I call Davidson's model of Psychological Explanations) in the following definition:

> A reason *r* explains an action *a* iff 1) there are descriptions under which *r* rationalizes *a* 2) there are descriptions under which *r* is nomically related to *a*, and 3) the descriptions of *r* and *a* under 2 are *logically independent* of the descriptions of *r* and *a* under 1 (this last condition reflects Hume's stricture).

7. Problems For Davidson's Model

Davidson's model of psychological explanations is an attempt to show that the same reason can both explain an action by rationalizing it under its intentional description, and cause it under a nomic description, thus eventuating in explanatory force. And we can now notice that which was implied above, that Davidson's model of psychological explanations is *almost* Davidson's theory of mental causation (in Davidson 1970/1980). Namely, the idea that a mental event can be said to enter into a causal relation with another event, mental or physical, only if there is a strict law which subsumes them both, under their physical descriptions. On the face of it, all we need to do to derive Davidson's model of mental causation from his explanatory model is replace the condition of logical independence in clause 3 with that of nomic independence. But on a closer look, there is also the question of whether the rationalization relation expressed in clause (1) is a necessary condition for mental causation. For if it is not, the question arises what implications this would have for the role of the mental/intentional in the production of behavior and other mental states.

It has been noted by a number of philosophers that Davidson's model of mental causation, and by extension, his model of rationalization, implies the doctrine of mental type-Epiphenomenalism (Type-E). This is the idea that mental events do not have causal powers in virtue of their falling under mental types, namely, *qua* mental (see Kim 1984b, Sosa 1984).[19] This seems to be implied from PAM and the principle of the nomological character of causality. Consequently, Kim has claimed that given these two principles, an event's

> causal powers are wholly determined by the physical description or characteristic that holds for it; for it is under its physical description that

[19]Type-Epiphenomenalism should be distinguished from Token-Epiphenomenalism, the idea that mental event-tokens can be caused by physical events but not cause anything.

> it may be subsumed under a causal law. And Davidson explicitly denies any possibility of a nomological connection between an event's mental description and its physical description that could bring the mental into the causal picture... Thus, Davidson fails to provide an account of psychophysical causation in which the mental *qua mental* has any real causal role to play. (Ibid., p. 267)

It seems that there is no better way to exhibit this point than in one of Davidson's own cases:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. (Davidson 1963/1980:79)

The problem which Davidson seems to have identified here is that a reason could rationalize an action, as well as cause it, and yet not do so "in the right way". What the case of the decomposed climber seems to show is that it cannot be sufficient for a reason to rationalize an action under its intentional description, and cause it under its physical description, if it cannot also be shown that the action was caused by the reason *in virtue of* its intentionality. We can see this perhaps more clearly with respect to a comparative case in which another climber (a vicious one) would have let go of the rope in order, or with the intention, to relieve his load. It would seem that the difference is that in the latter case, the mental state plays a causal role in virtue of its *instantiation* of some causally relevant property, which is absent in the first case. And this, one might expect, is a causally relevant *mental* property, *pace* Davidson's own theory of mental causation.

8. Two Theories of Explanation

It seems that the route I have been taking up to this point refuses to lead us away, as some would like to see, from the question of the causal powers of mental properties as such, as opposed to the causal powers of mental events in virtue of their having some physical property. Which, by the Humean conception of causation, implies that there must be mental/intentional laws. Thus, in my opinion, the problem that Davidson is facing is not one about *explanatory* force, since, as I claimed above, explanatory overdetermination is built into the D-N model of scientific explanation, and so is not specific to rationalizations. The problem, rather, is whether reasons (as reasons) *can be* causes because if the

answer is in the negative, then there is no problem at all with explanatory overdetermination. This is a point which Kim makes with respect to Kuhn's incommensurability thesis (Kim 1989, Kuhn 1962), that only the idea that there could be at most one actual cause can make two adequate explanatory schemes incompatible (this point also underlies Kim's "explanatory exclusion principle"). Thus it seems that Davidson's problem of explanatory force and rational overdetermination in fact *presupposes* that reasons can be causes, but fails to supply a model that makes them causes as such. But could he, given his adoption of PAM?

According to Loewer and LePore (1987), there is a way to make reasons into causes even without invoking psychological laws. Part of their argument relies on the claim that the charge of Type-E against Davidson confuses two notions of causal relevance, which they define as follows: Properties F and G are causally relevant$_1$ to c causing e just in case c's having F and e's having G makes it the case that c causes e. Relevant$_2$ is defined as the case where c's being F brings it about that e is G. Then they claim that Davidson's theory can make 'mind matter' because it entails that mental properties are irrelevant in the first sense though not in the second. Thus, what we have here is not a conception of how certain thoughts cause an action by being the thoughts they are, but how *the fact* that a physical token happens to instantiate these thoughts brings about the instantiation of the action by another physical token. Perhaps a better understanding of what is involved here, as well as some important insights, will be gained if we compare L&L's two ways of causal relevance to Cummins' two theories of explanation - Transition theory and Property theory (in Cummins 1983).

A transition theory specifies the conditions sufficient for a system to undergo a change of state, and it does so by reference to some (not necessarily Hempelian) model of nomic subsumption. This seems to be what L&L have in mind in the first kind of causal relevance of a property. Then, like L&L, Cummins thinks that there is another strategy of explanation which is concerned not with transitions but with instantiations. According to Cummins, a Property Theory aims to explain not how a system changes from one state to another, but what it takes for it to instantiate a property, given its other instantiations at that time. As can be seen, this is a kind of a functionalist, or at least, second-order, model of instantiation, to which I shall return in the following chapters. But right now we can see that such a model might lend itself to some account of the causal relevance of mental properties, and without relying on their nomic powers. That is because, according to Cummins' Property Theory, a mental property can become causally relevant if its instantiation by a system (e.g., an agent) would be sufficient to bring about the instantiation by the system of another property, mental or physical. As can be seen, this is just L&L's second conception of causal relevance I discussed above.

Cummins' Property Theory does not apply only to the mental, but to any case where the instantiation of some properties are taken to bring about events in the world. Thus, if such a model succeeded, it would seem that we could have an account not only of the explanatory relevance of singular mentalistic explanations, but also of other everyday macro explanations. Recall Davidson's example above with 'the event reported on page 5 of Tuesday's *Times caused* the event reported on page 13 of Wednesday's *Tribune'*. There I agreed with Davidson that we have no use for such a causal statement. But how about 'The hurricane caused the catastrophe'? This looks like a good explanatory causal statement, even if we agree with Davidson that there are no (scientific)[20] laws which would use such terminology. But then, perhaps all there is to mentalistic explanations is what is also true for any explanation in macro-terms (terms for macro events), as is expressed by Cummins' Property Theory? This seems to be Antony's idea:

> If we want to understand a state transition in a complex system, we may need to learn first through a property theory how the antecedent and consequent states are realized in that system, and only then look for transition theories to subsume the lower-level events under causal generalizations. This is *obviously* how it's going to work for hurricanes and disasters (1989:172, my italics).

However I want to express my scepticism that this explanatory strategy is so obvious, and while I'm at it, I would also like to show some more general doubt with respect to the attempt to explain the causal relevance of mental properties via the Property Theory. This will also give me the opportunity to say what direction I think that a theory of mental/intentional causation should take.

To begin, I doubt it is the case that when we understand a statement such as that the hurricane caused the disaster, we do so through an analysis of a "functional problem" - a problem about what it takes for a system to instantiate a property such as being a hurricane. And if that is the case with hurricanes, windows, chairs, parties and the like, it is even more so with the case of mentalistic explanations. For do any of us have any idea about the instantiation base of such states? Or, do we need to know of such an instantiation base to understand and issue mentalistic explanations?

Then, there is also the problem of whether functional terms can be explanatory in the first place. Compare: 'the mouse was trapped by the mouse-trap because it catches mice', or 'the window was broken by the stone because

---

[20]I want to keep the option to claim that not all laws have to be scientific laws.

a stone is the sort of thing which breaks windows'. Isn't it simpler to assume that we understand that stones break windows because *if* you throw a stone at a window, *then* it would most likely break (unless the glass is strengthened, or thick, or you didn't throw it hard enough, etc)? And yes, this probably has to do with the stone being heavy, having a certain kinetic energy, etc., but *how* heavy, and *how much* energy?

I would argue that these kinds of speculations are not what makes our macro explanations successful, and hence I would dismiss the claim that they presuppose a Property Theory. More likely, I think that answering such explanatory problems presupposes reference to counterfactuals with dependency conditions, counterfactuals which involve projectible[21] properties, including such "non-scientific" properties as *stone, hurricane, mousetrap, party,* and indeed, mental states, but not *either a stone or a mouse-trap, a stone and 2+2=4, undetached stone parts,* etc. , properties to which, arguably, creatures with mentality have some special access. My claim is that it is the special capacity of minds to detect such properties and their modal relations which grounds our ability to make useful predictions and explanations in their terms even though these properties are not what we regularly take to be nomic properties.

Well, in fact, it will be my claim later on in this dissertation that such properties *are* nomic, in some special way, and that the same goes for mental properties. In this I agree with Antony that explanations in terms of macro-terms and in terms of mentalistic terms are on a par, but not because they presuppose knowledge of something like Cummins' Property Theory. Still, I accept that Cummins' Property Theory can be part of a certain *metaphysical* story about the relations of macro properties to their realization base, including the case of mental causation. I will finish this chapter by a preliminary explanation of what I have in mind here, and also how I see the solution to the problem of whether reasons can be causes.

9. The Theory of Modal-Disjunctive Properties (I)

I want to begin this section with a look at what we have seen so far concerning the issue of intentional causation and where the dialectics of the discussion has brought us. I began the discussion in this chapter with the thesis that psychological explanations and predictions are a matter of subsuming mental

[21]According to Goodman, a projectible property or predicate is one with respect to which we are justified in expecting unexamined objects to resemble examined ones (1955, ch. IV).

states and events under intentional laws. We saw that underlying this claim was a certain assumption about the place and role of content for psychological processes. I put this idea more specifically in terms of the Fodorian thesis that the logical relations among the propositional objects of mental states constrain their causal relations to one another. This I called the isomorphism thesis (IT). At this point, what seems to be a conceptual argument against this thesis was presented. The claim was that mental states cannot enter into causal relations which "mirror" their conceptual relations. As we saw, this claim goes back to Hume's stricture - the constraint that only logically independent events can enter into causal relations. The application of Hume's stricture to the case of the mental seemed to imply that rationalization cannot be a species of causal explanation, as the anti-causalists concluded. This conclusion was objected to by Davidson on the ground that reasons which are not causes lack explanatory force. Instead, Davidson suggested a model of psychological explanation which is a combination of two models: the logical model of the normativists, and the nomic model of Hempel. According to Davidson's combined model, a reason explains an action just in case there are descriptions under which it rationalizes it, and distinct descriptions which subsume both under a causal law. The first component is supposed to supply the intentionality of the act, the second its explanatory (causal) force.

Now in effect, when put in this formal mode, Davidson's solution seems to raise the suspicion that it relies in the questionable analytic/synthetic distinction[22]. For I see no other way to interpret the claim that a principled distinction has to be made between sentences true in virtue of their logical entailments and sentences true in virtue of their instances. And as quite a few philosophers, including Davidson (for example, in Davidson 1974), have been convinced by Quine's argument against that distinction (in Quine 1953b), this surely should give us pause.

However, once we turn to the material mode, there lurks, in my view, an important insight in Davidson's theory. It seems that rather than a questionable theory of description types of events at distinct "levels of explanation" (of the sorts of rationalizing, explanatory, nomic, referring descriptions, etc.), there is implicit in Davidson's account a dual theory of mental explanation according to which there are *two* ways to explain why a mental event, such as an action, took place, ways which, and this is the important point, are not completely independent of each other. That is, it seems to me that the moral we should take

---

[22]The analytic/synthetic distinction is the claim that there is a principled distinction between those sentences true (or false) because of their meaning only and those sentences which are true (false) both because of what they mean and because of their empirical content, that is, the way the world is.

from Davidson's theory is that in explaining an action by reference to reasons we give the purpose of the action but not *how* it came about, whereas giving the physical description, the *mechanism* as some call it, would supply the explanatory force, but not the *why;* but it is only when both the *why* and the *how* were supplied that the action would have been *causally* explained. I want to give now a rough sketch of what I think is involved in such an amalgam of event determination, drawing, as I said above, on Cummins' Property Theory.

$$\begin{array}{ccc} F & \rightarrow & G \\ \downarrow & & \uparrow \\ Mf & \rightarrow & Mg \end{array}$$

Figure 1.1: Schematic account of the implementation of "higher-level" processes by "lower-level" mechanisms.

Figure 1.1 is a rough depiction of what Property Theory takes to be involved in the relation between two explanatory nexuses. In this schema, F and G stand for the "higher-level" properties, such as mental properties, and Mf and Mg as their physical realizations. In Davidsonian terms, Mf→Mg would be the process which gives the explanatory force to F→G, and that in virtue of its being subsumed by the appropriate physical law (not shown here).[23] F→G, on the other hand, explains *why* the "lower-level" process took place in the first place, that is, rather than some other process. Taking 'F' to stand for the reasons and 'G' for the action, the idea then is that the instantiation of the reasons would be sufficient (given certain conditions) for the instantiation of the action via, first, the sufficiency of the reasons for the instantiation of a realizer Mf - in some sense of sufficiency yet to be explicated - and second, the sufficiency of Mf, a causal sufficiency this time, for the instantiation of the other physiological state represented by Mg. It is only then that we arrive at the action via the presumed sufficiency of Mg for its instantiation. Now, to be sure, endorsing such a model for the implementation of mental processes depends on our acceptance of some systematic relations between the two levels, the one represented by F→G and the one represented by Mf→Mg, and this would seem to go against Davidson's principle of the anomalism of the mental. This, surely, should not mean that we need be deterred by this, but in fact, it seems that even Davidsonians can rest assured here that this model is compatible with PAM. That is because while in

[23]Someone might object that what we have here are not really two explanations of one event but of *two* distinct events - G and Mg. But that is wrong, I would claim, since it is the *action* we are trying to explain by *both* explanatory strategies.

PAM Davidson has expressed his view that there cannot be (strict) causal laws which subsume the mental as such, the kind of laws we suppose underlie the "vertical" relations in Property Theory are not causal laws but bridge-laws.[24] And indeed, Davidson has conceded in a number of places (most recently in his 1993 paper, pp. 8-9) that he has he no objection to there being *non-strict* psycho-physical and psychological laws. Apparently, his objection is only to the idea that non-strict laws can be causal laws, whereas, as I said, bridge-laws are not causal laws. So unless Davidson has an argument specifically against the possibility of non-strict laws being bridge-laws, I cannot see PAM to be incompatible with Property Theory.

Given this bit of Davidsonian scholarship, we can now return to our model and look more closely at the array of dependency relations that it implies between the intentional process and its physical realization. I claim that what we are looking for in this model of 'implementation' is a sort of a dependency relation between the intentional and mechanistic aspects, but one which is conditional rather than reductive. Let me will explain. One way to view the relation between the two explanantia is by reference to some nomic dependency relation, say by reference to bridge-laws connecting the terms of the higher-level explanation (F and G) with those of the lower-level explanation (Mf and Mg).[25] In which case the transition from F to G is in fact reduced to that from Mf to Mg, such that the law which governs the process is a lower level law, a mechanistic law. But of course, in a case like this we no longer have *two* explanations of the action but only one, even if we stop short of *identifying* the two levels, making them only nomically dependent. Indeed, that model might be perhaps explanatory for *theoretical* reductions, in which the micro-explanation gives a more revealing account than what the macro explanation possibly can. A classical case in point is the reductive explanation of the expansion of gas when heated by reference to the increased kinetic energy of the gas molecules. But this model would just bring back the Type-E I rejected above, for this is surely not a case where the higher level properties make their own independent contribution to the process.

The alternative to reduction is supervenience, which is an asymmetric dependency relation. A weak version of supervenience with respect to the

---

[24] The notion of a 'bridge-law' is due to E. Nagel 1950.

[25] The locus classicus for this type of reduction is E. Nagel 1961. For someone who claims that all reductions through bridge-laws must be identities, see Sklar 1967.

relations between the mental and the physical we find in Davidson:[26]

> ...mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but different in some mental respect, or that an object cannot alter in some mental respect, without altering in some physical respect (1970/1980:214).

The idea here is the following. For any two property-types A and B (such as mental and physical properties), A supervenes on B just in case there is no change in A without some change in B, or alternatively, a change in A *implies* a change in B. This means that for each F→G, there must be some lower level physical mechanism to realize it, or in other words, sufficient for its instantiation. But then, each and every such physical mechanism which realizes a higher-level mental process would be dependent on that mental property, because if B is sufficient for A, then A is necessary for B. The result of this is that while the supervening properties depend on there being *any* property of type B to get instantiated (a physical property), any *particular* physical property instantiated would depend on some particular mental property.[27]

If we apply this model of dependence to our combined explanatory model, making F and G supervene on Mf and Mg respectively, it would follow that there could not be a change from F to G (F→G) without *some* change from Mf to Mg (Mf→Mg), but that any particular $(Mf \rightarrow Mg)_i$ would be sufficient for G and hence depend on G. But this is not what we want, because we want to be able to account for cases where the physical realization can take place *without* the action, as with Davidson's climber case I discussed above. To wit, if the intentional action of the vicious climber was to be described as 'letting his partner fall', then we want to allow a case where someone can make the same physical

---

[26]The weak version constrains the dependency relation only within a world. In that sense, the "cannot" here is of the material conditional and hence lacks modal force. The strong version of supervenience implies psychophysical laws, hence not something Davidson would advance. On the differences between weak and strong supervenience, see Kim 1993, especially chapter 4.

[27]We can compare this to the case of causation: We say that if c causes e that is because it is sufficient for it. But then in what sense does the effect depend on the cause and not the other way around? Only in the sense that (*ceteris paribus*), if a cause of that *type* hadn't occurred, the effect would not have occurred. But each particular effect does not depend on its particular cause but vice versa.

movements without performing that action, as with the case of the decomposed climber. On the other hand, we do accept that the action would not have been performed unless the physical mechanism was there to take it through.

The solution is to make the dependency between the two levels conditional: We say that the relation between the two levels has to be such that the intentional process would not have taken place unless the physical did, but that the physical is not a sufficient condition for the intentional, or at least not the physical as so far described. Let's turn to see how this works in the example of the climber who let his partner fall off the cliff since he thought that otherwise they might both tumble to their death. According to my revised version of the Property Theory, the idea is that in every possible world in which the particular reasons the climber had (in that particular situation) issued in that intended action, there was also some physical mechanism which eventuated in his loosening of the grip on the rope. But then we assume (or just postulate), that there are possible worlds in which the mechanism leading to his loosening the grip on the rope was the same but still the action was different - a case in point is that of the decomposed climber. In fact, we can see the decomposed climber as a counterfactual Twin of the vicious climber in which all the physical facts are fixed but the action, the intentional property instantiated by the behavior, is different. And to be sure, this is a different dependency relation than that of supervenience, since here the "lower-level" process is not a sufficient condition for the "higher-level" process.

But now the question is, how did the change in the "higher-level" process come about, that is, the change between the vicious and decomposed climber, without a change in the physical realization? My answer is that it comes about just in case a different physical property, one which is more "fine-grained" than the mechanisms I just discussed, is instantiated at the Mg position. This is an issue I'm going to develop more fully in later chapters so I will be brief here. The idea is that there are properties, in fact, intentional properties, which are constructed from certain disjunctions of physical characteristics. For the time being, we might call them 'access properties' because they are just so many ways that we, intentional systems, get access to the world. Thus, for example, we might presume that there are many and different ways for one to engage in letting one's partner fall-off a cliff: by loosening the grip on the rope holding him, by cutting the rope, by shaking it wildly back and forth, and so on. And we may assume that all these ways could connect with the very same reasons that the vicious climber had for letting his partner fall off. Then, alternatively, there are also all those different ways to respond to troubling thoughts: biting one's lips, making facial gestures, and...yes, relaxing one's hand muscles, the very hand holding the rope to which one's partner is clinging to dear life. It was the misfortune of the decomposed climber (and his partner) that he reacted in this latter manner to his troubling thoughts, rather than in any of the other ways. And

it was also to the confusion of the police investigators that the two properties, that of letting your fellow-climber fall to his death, and that of reacting frantically to some unsettling thoughts, overlapped at that physical juncture. The reasons were indeed the same, the mechanism, that is, the physical process, was the same, but the outcome, from the intentional point of view, was different. And to be sure, part of what makes intentional states what they are is that they are sometimes not satisfied (for more on this, see chapter Six). Thus, in a sense, we can see the case of the decomposed climber as the *error-case* of the vicious climber - of what happens when things don't go as planed, although still with the same physical result.[28]

Here we can see most clearly the dependency of the physical on the mental-intentional. For the physical process going from Mf to Mg to count as implementing an intentional process, as is the case with the vicious climber, it is necessary that it sometimes, that is, in some possible worlds, results in an action which is not the intended action. Or less question beggingly, it is required that there would be conditions under which the same physical process, starting with the same reasons, eventuate in a behavior with a different intentional description (or even no intentional description at all, as when it is only the blood pressure of the climber which climbed up). And what is more, it is also required that the case of the decomposed climber *depend* on that of the vicious one in that his unsettling thoughts would never have eventuated in the unfortunate response were not the mechanisms which enable vicious climbers to get what they want already in place.

Thus we can see that the physical implementation of the process from reasons to action in the case of the vicious climber would not have taken place *as such* were it not for the possible instantiation of the other process, the one with the more innocent and distinct intentional description. But on the other hand, as I have said already, no intentional process could get off the ground were it not for some mechanism to take it thorough.

To sum up, then, the idea being put here is that *pace* the supervenience thesis that Davidson advanced above (and to which I shall return in later chapters), it is possible for the same instantiation of a physical property to realize a distinct intentional property since there is more to the realization of intentional properties than their physical coextension. That is, my claim is that when we go to consider the physical realizer of a certain intentional state, we need to go

---

[28]Imagine a novice assassin who aims at his target, but at the last second is so startled by the thought of shooting someone that his finger presses the trigger and he kills the man. This is not someone you are going to call a professional assassin.

beyond its actual realizer, to all of its nomologically possible realizers as well.[29] This is because the particular physical property instantiated might belong to a distinct disjunctive realization base which might just happen to overlap with it, as in the example I reviewed. It is because of this that the counterfactuals matter more than the actual realization, something which Davidson, surely, would find objectionable. But isn't it usually the case that in cases of indeterminacy, as that relating to the correct intentional description of the climber's behavior, we would most likely ask: what would he had done had his loosening of the rope not eventuated in the fall of his fellow climber, say, in case the rope got somehow stuck in a hook in his pants? It is on the answer to this question that the appropriate intentional attribution of the climber's behavior depends. If we conclude that our climber would have immediately pulled on the rope and then anchored it to something secure, then he is the decomposed climber and the physical movement corresponding to the loosening of the grip is part of a disjunctive array of other possible responses to frantic thoughts. But had we concluded (say by reference to some further evidence) that he would have released the rope in any case, thus letting his "buddy" fall, then he is the vicious climber and the same physical movement and mechanism are now part of a different disjunction which corresponds to all those ways a climber might attempt to get rid of his partner in the said situation.

Some might wonder at this point how the physical system which is the climber can make sure that it would be one action that would come about rather than another, if the physical process, we assume, is the same in both situations. That is, the question is how can the climber make sure that he is, say, the decomposed climber rather than the vicious one if both are underlay by the same physical mechanism. In the words of Fodor (1968, p. 637), the question might be put as "How does one produce behavior of that kind?" In contrast to the mechanistic model of psychological explanations which attempts to analyze "higher-level" process into lower level and less intelligent ones, where *no* answer could be given to this question (*pace* Fodor, ibid.), a D-N model would not need to. For as the nomic model goes, it is the subsuming generalizations which matter, and those, I claimed, rely on the counterfactuals for their confirmation. In the climber case, we might presume that the conditional 'people who are in climbing situations and believe that their climbing partner is too heavy tend, *ceteris paribus*, to let them fall to their death' is, luckily, rarely true, which shows

[29]This might sound a bit like Kim's theory which collapses supervenience to a reduction to the disjunction of all possible realizers. But as will become clear in chapters Six and Seven, my theory differs from that of Kim in some crucial respects. The upshot will be what I take the scope of the nomologically possible to be.

that the vicious climber is an aberration. That is, in a case like this, we should see the behavior of the climber more truly described as the action of 'accidentally letting one's partner fall'. On the other hand, the vicious climber case will need another conditional to subsume it, as that 'psychopaths who are in climbing conditions...etc'. But this indeed shows that we need the modality of laws to make sense of the intentional description.

One issue that is still left open in this nomic account of intentional causation is that concerning Hume's stricture I discussed. For independently of Davidson's other motivations for PAM, Hume's stricture should be respected, I would argue, even by those who either do not think that causal laws need be strict, or those who do not buy into the notion that the intentional domain is determined by its unique proprietary set of constitutive conditions.

My claim is that our asymmetric dependency model for intentional causation can show a way to avoid the problems with Hume's stricture. We recall that what Hume's stricture requires, in effect, from a theory of intentional causation is to keep the states which enter into causal relations independent of each other. That is, the problem is not whether the states are logically implicated, since that is actually the case with any two events which enter into causal relations. That is, since for any event c causing event e, c can always be given the description 'the cause of e' and thus make the sentence 'c caused e' into a tautology. And so we accept Davidson's point that not just any way to describe events is a legitimate way to subsume them under laws. But what I think Davidson and the other normativists have missed is that events can be described in ways which are both logically and nomically implicated, but still the events would be independent events. This point is exhibited nicely in our model of asymmetric dependency condition for mental action. What we saw above was that although the reason entails, or justifies, the action, there are still worlds in which the reason is instantiated but the action, say that of letting one's partner fall to his death, is not. We achieve this by making the state which nomically realizes the action still not sufficient for it, as cases of error could, and in fact *must* occur, for this physical state to be considered as a realizer of an intentional state.

As I said, I will return to this modified Cummins' model later in the dissertation, but from now on my main concern will be rather with the theory of properties it introduces. In the next chapter I will try to show that this theory can be useful to strengthen other weaknesses in the theory of intentional causation, and in its underlying thesis of isomorphism between the mental and the causal. Some of these difficulties have to do with the issue of "grain", others with the issues of wide versus narrow content, and some with the theory of content as such. All these will be discussed in the following chapters.

Chapter Two

The Isomorphism Thesis (IT)

In the last chapter we saw how certain assumptions about the nature of psychological explanations and predictions (PEPs) were tied to some controversial views about the nature of mental causation. In particular, we saw that the intentional nature of the Folk PEPs, plus some assumptions about their modality (that they support counterfactuals, are supported by their instances, etc.) were translated by intentional realists into a conception of causation whereby mental states have their causal powers in virtue of their content. However, we also saw that accepting the vocabulary and generalizations of the Folk, if only as a starting point for a more mature science, put intentional realists in a precarious position. The reason was that in addition to their nomic character, PEPs are assumed to be constrained by the semantic relations that obtain between the states they subsume. I called this sort of constraint, where the causal is supposed to "respect" the semantic, the thesis of isomorphism (or IT for short), and noted that it seems to be in tension with a general stricture, introduced by David Hume, that causes should be logically independent of their effects.

Although a tentative solution to the stricture problem was introduced at the end of chapter one, it is clear that a more comprehensive treatment of this issue can come only after the isomorphic thesis itself has been more specifically characterized. This will be to a large extent the project of this chapter, where IT will be considered from several points of view. First, I will try to give a more precise formulation of the thesis via an explication of the technical notion of isomorphism. Then I shall discuss one famous (or infamous to some) attempt to substantiate IT through a metaphysical reduction of content to causal role. This is conceptual role or functional role semantic theory (henceforth, CRT).

For some time, CRT showed itself to be the most promising and developed attempt to give a reductive analysis of the idea that the semantic relations between the objects of the attitudes are aligned with their causal roles. Surely, were it to succeed, it would amount to a major contribution to intentional realist theories such as RTM. However, it did not take long for CRT to come under sharp criticism, the main brunt of which was directed to its inherent verificationism. For given Quine's repudiation of the analytic/synthetic distinction, verificationism plus the non A/S distinction leads to meaning-holism and then to meaning eliminativism.

After presenting these arguments against CRT as well as others, I shall

review attempts to rescue the theory by reducing content to functional role directly, that is, not via inferential role. However, I will claim that these attempts reintroduce verificationism in the form of ideal verificationism. I will end the chapter by suggesting that a new notion of content is required, one not reducible in any way to inferential role, and which is also more "fine-grained" than that which CRT can give us. This is informational content, a specific version of functional role semantics which will be discussed later on in the thesis.

1. Isomorphism

As a mathematical notion, isomorphism is a relation that obtains between two mathematical objects or domains - such as the natural numbers and the fractions - so as to preserve, in both directions, some intended structural aspect. This is accomplished through *mapping* each structural element in one domain with a 'corresponding' element in the other domain and vice versa. Such notion of 'correspondence' prompted some people to give a definition of isomorphism in something like the following:[30]
Two structures S1 and S2 are isomorphic iff there is a mapping relation between them such that:

1) For every object in S1 there is exactly one corresponding object in S2 and vice versa.
2) For every property defined in S1 there is exactly one corresponding property defined in S2 and vice versa.
3) whenever a relation defined in S1 holds of *n*-tuple of objects in S1, the corresponding relation in S2 holds of the corresponding *n*-tuple of objects in S2 and vice versa.

The problem with this attempt to define isomorphism, though, is that it would lack substance unless the notion of 'correspondence' could itself be independently defined. Otherwise, any two structures with the same cardinality could be said to be isomorphic, depending on one's choice of a scheme of correspondence.

The importance of the correspondence issue can be illustrated when we try to utilize the definition of isomorphism to substantiate IT. To wit, we recall that IT requires that the causal relations between mental states follow the same pattern of semantic relations, to be discussed below, between their propositional objects. But to be sure, not just any correspondence of patterns would suffice here. That is, we can foresee a situation where one pattern is inverted such that although any semantic relation between the attitudes still corresponds to some causal relation and vice versa, they do not correspond "in the right way". The right way would be when the causal structure would represent the *relevant* semantic properties of the attitudes, on their inter-relations.

[30]Adopted in part from Cummins 1996:96.

Here is an analogy (due partly to Cummins 2000). The structural relations between the causal properties of the attitudes and their semantic properties can be seen as something like the structural relation obtaining between a map, say of the city of Chicago, and its target domain. Indeed, the whole point of a map, or a model, is that of exploiting some of the structural relations between the objects of the map (or model) to represent structural relations in reality; for example, of using the comparative length and orientation of lines on the map to represent distances and orientation of locations in the city. But to be sure, not every structural aspect of the map is representational, even those which might happen to convey information. For example, surely it is the case that any such map bears a structural correspondence to other maps of the city of Chicago. Still, we would not think that it represents those other maps. Of course, this is not to say that one cannot *use* it for that purpose, but we probably then say that this is not the right use, in the same way that one could improperly use the map by holding it upside down.

The idea that selecting the proper isomorphic relation between two domains has to do with proper or improper use gains support also from this example by Dretske (1986:23):

> A sensitive spring-operated scale, calibrated in fractions of a gram, is designed and used to determine the weight of very small objects. Unknown to both designers and users, the instrument is a sensitive indicator of altitude. By registering a reduced weight for things as altitude increases...the instrument *could* be used as a crude altimeter if the user attached a standard weight and noted the instrument's variable registration as altitude changed. Suppose, now, that under normal use in the laboratory the instrument malfunctions and registers 0.98 g. for an object weighing 1 g. Is it misrepresenting the *weight* of the object? Is it misrepresenting the *altitude* of the object? What does the reading of 0.98 g., mean?

According to Dretske, the reading of 0.98 g means a (incorrect) reading of the object's weight since this is how the device was designed to be used, that is, by its intended function. Similarly, Cummins has argued that a speech parser processes information about the phrase structure of the linguistic input rather than, say, its pitch, because this is its function, a natural function this time (2000).

To be sure, both Dretske and Cummins are discussing the problem of (mis)representation rather than that of IT, but I think that the issues are pretty much the same. That is because in both cases the task before us is that of explaining how the causal properties of physical structures, such as mental or linguistic representations, can "mirror" semantic relations of some intended domain. Because of that, I would claim that IT shows itself to be a theory in the

family of representational theories, theories which are concerned with the question of what makes certain causal structures have the representational properties, or content, they have. We saw that as per Dretske and Cummins, how representations are *used* is at least part of the answer. I shall now turn to a theory for which the way in which how a representation is used is a *large* part of the answer about meaning, a theory which seems, indeed, to be tailor-made for IT.

2. Conceptual Role Theory (CRT)

The question of what makes a representation, or any meaningful expression, meaningful, is a metaphysical question, and hence requires a metaphysical solution. In other words, it requires a reductive theory which could give in non-semantic and non-intentional terms the conditions sufficient (and perhaps necessary) for a structure to have semantic properties. This is how proponents of CRT see it (Block 1986). Thus, CRT is a metaphysical thesis about the constitutive conditions on mental content and linguistic meaning.[31] In the spirit of Wittgenstein's dictum that meaning is use (1953), CRT takes the content of a representation to consist in its functional role "in the cognitive life of the agent". This includes perception, thought, and decision making. We are going to see that there are in principle three kinds of functional role theories, which correspond to three ways in which meaning can be determined by inferential role. One type of CRT is called two-aspect or two-factor theory, according to which meaning is divided into a narrow and "wide" aspect (Block, 1986; Field, 1977; McGinn, 1982). As opposed to two-factor theories, there have been suggested one-factor theories like that of Harman (1987), according to which conceptual role can also reach out to reference (also called "long-arm" CRT).[32] A third version of CRT is one in which narrow meaning is taken to *determine* reference (perhaps given a context. See Fodor 1980, 1981, 1985, 1991).I shall review below each of these theories briefly, but first some general remarks.

One of the strongest motivations for the development of CRT in

---

[31]Although CRT sees itself as both a theory of mental content and a theory of linguistic meaning, I shall ignore in what follows the linguistic issue.

[32]Another example of a "long-arm" theory is that of J.J.C. Smart. See his 'topic neutral' definition of a yellow after image as *"there is something going on which is like what is going on when* I have my eyes open, am awake, and there is an *orange* illuminated in good light in front of me" (1959, second italics mine).

philosophy in the early 80's was the attempt to provide a reductive theory of meaning which would be able to accommodate the bifurcation of meaning into two meaning 'factors': "wide" and "narrow" (Putnam 1975).[33] To wit, this is the idea that meaning is divided into one factor which is referential (or truth-conditional), and then another factor of meaning which can be common across reference. As we saw with the case of Putnam's Twins (in the introduction), the referential aspect of meaning differed between the twins (H2O for the earthlings and XYZ for twin-earthlings) whereas the descriptive aspect, also associated with their behavior, was presumed to be the same ('the liquid, potable, transparent stuff, etc'). Proponents of CRT thought to construe the narrow aspect of meaning (also corresponding to Putnam's stereotypes) in terms of functional or inferential role, and the wide aspect of meaning in terms of a truth conditional theory.

With respect to the referential component, the favorite candidate is usually Davidson's theory of meaning based on a theory of truth for a language (1984).[34] That is because Davidson showed a way how to construe meanings in terms of truth conditions, issuing in the famous extensional biconditionals of Tarski's convention T (in Tarski 1944):

T: 'S' is true in L iff P

Here 'S' stands for a structural description of a sentence in a language L, and P for a sentence in the used (meta)language which is supposed to be giving the meaning of 'S' in terms of its truth-conditions, thus:
"Water is wet" is true in English iff H2O is wet,
and
"Water is wet" is true in Twin-English iff XYZ is wet.[35]

---

[33]This is not to say that CRT started as a response to Putnam's Twins. In the non-philosophical circles, most notably in AI, versions of CRT, under the name of 'Procedural (or computational) semantics', were known to exist already in the early to mid-seventies. For a representative examples see Johnson-Laird, 1977; Miller and Johnson-Laird 1976, Winograd 1971. For a criticism of procedural semantics on the same lines to be introduced below, see Fodor 1978.

[34]In fact not quite Davidson, since Davidson's project (at least the one expressed in the collection of articles in Davidson 1984) was of how to construct theories of *interpretation* for a language rather than strictly a theory of meaning.

[35]The idea here is that coextensive substitutions are allowed in the used sentence but only to an extent. For example, even a truth-conditional theorists would not want to construe the meaning of 'water' in terms of any coextensive

The narrow aspect of meaning which CRT proponents advance is its so called "opaque" meaning, and has to do with how agents *represent* things to themselves. This is also taken to be the aspect which is more relevant to the causation of behavior and other states.[36] Hence CRT's underlying idea of narrow meaning as functional role.

It is important, though, to distinguish this theory from a similar theory according to which narrow meaning is the Fregean component of meaning, its Sense, or mode of determination of reference (this theory will be discussed more fully in the next chapter). Fregean theories of mental content[37] take one component of meaning to be its reference, or truth-conditions, just like the referential theory I mentioned above, and another component to consist in the mode of referring to these conditions. This mode of reference consists just of those inferential relations which are relevant for reference determination. For example, the Sense of 'Bachelor' would consist of 'unmarried adult male' since 'unmarried adult male' refers to anything which 'Bachelor' does. It would not consist of other inferences that people tend to draw from 'Bachelor', say 'living a happy untroubled life', because such descriptions do not reliably share a reference with 'bachelor'.

Fregean theories are not suitable for CRT's conception of narrow content because they are representational (see Devitt 1996), and hence do not satisfy Putnam's dictum that '(representational) meaning is not in the head' as exemplified in the Twin cases. In contrast, CRT takes the narrow component of

---

term, say 'Granny's favorite concoction'. But so far no principled distinction between allowed and not allowed substitutions has come forward, and I suspect none will.

[36]CRT proponents acknowledge that wide meaning is also relevant to the explanation and prediction of behavior. For we explain Jane's fleeing behavior by reference to tigers and not to their phenomenal descriptions. Indeed, people can represent tigers to themselves in all sorts of ways ('a funny-looking zebra') and still exhibit the same behavior. The six million dollars question is, of course, what explains the *success* of wide explanatory strategies which are commonly used by the Folk, and according to some (Burge 1986, Wilson 1995) by scientists as well. This issue will be the topic of chapter Five.

[37]I know that this sounds somewhat confusing given Frege's Platonism, but I refer the defenders of Frege to Putnam 1975 in which this very idea is exploited in the argument. In any case, the a Fregean kind of mental content is one where the descriptive content of a mental state determines its reference. Whether *this* is consistent with Frege's own view of Senses is a question for Fregean scholarship which I will touch upon in the next chapter.

meaning to be non-representational, leaving the referential part to be determined wholly by the truth-conditional aspect. In this way, molecular Twins can be identical in their narrow aspect of meaning and distinct in their truth-conditional aspect since, to repeat, the first does not determine the second. This also means that the range of inferential relations that enter into the determination of narrow meaning can be much wider, since it is not constrained by referential considerations (that is, it can include inductive or synthetic inferences).

Now, if narrow content does not determine reference and truth, then one might wonder whether it should be called meaning at all. Because of this, there were those who suggested that narrow meaning should at least determine some "analogue" of truth, perhaps warrant (Boghossian 1993). That is, it might be said that people who infer from 'Elmer is a bachelor' to 'Elmer lives a happy untroubled life' are warranted in doing so (perhaps by communal standards), although no determination of reference can come out of this. We shall also see that this non-representational or non-truth-conditional aspect of CRT is connected to its verificationism, but I defer that until later. In any case, this non-representational aspect of CRT should raise some doubts as to its ability to substantiate IT, that is, as a theory that attempts to explain truth-preserving processes (rather than truth*-preserving processes, where truth* is 'truth by warrant'). But then, it might be claimed that one cannot be prejudiced as to what notion of truth we should be working with so I will leave it at that. Be this as it may, there are other aspects of CRT which are independent of its non-representational version that are still relevant to our discussion. Thus I will continue and raise this issue only when it matters.

CRT is a functionalist theory. It identifies the semantic[38] properties of certain structures by reference to their functional properties: properties which *state-tokens* have in virtue of their causal dispositions (Loar 1981). Because of this, it can be said that CRT solves the problem of IT by fiat. For the idea is that the causal roles of mental states "mirror" their semantic relations because these causal roles *constitute* those semantic relations. Thus, for instance, conjunctive thoughts tend to cause thoughts about their conjuncts because that is part of their causal roles. This is just how they are defined.

But a question arises. Although some of the roles of mental states can be described in terms of inferential relations - deductive, inductive, decision making, etc - surely that cannot be the case with all of them. Some, such as those

[38]The sense of 'semantic' that I use here is the more general sense of it as 'relating to meaning'. Thus, on this conception both the truth-conditional aspect and the functional aspect of meaning are semantic. This use of the term differs from that by Tarski and his followers for whom semantic has essentially to do with the relation between language/mind and world things and affairs.

pertaining to input and output, are merely causal. Even more so, that would be the case under Harman's conception of CRT as involving perceptual and behavioral relations to objects in the world. But if this is the case, then one might wonder what justification CRT has for the claim that such causal roles can be said to *define* the semantic properties of these states. After all, there are all sorts of causal relations in the universe, most of which have no semantic significance (*pace* Spinoza). So CRT needs to give us some demarcation between those causal processes which are meaning (or content) constitutive, and those which are not (cf. Fodor 1987:77). What we need is a theory that connects the causal-roles of mental states with their semantic properties in a non-question begging way; that is, without (direct) reference to their intentional properties. The trick would be to construct a functional theory for mental states which somehow *presupposes* the intentionality of these states. I will say first a few things about functional theories, and then how we might satisfy the latter constraint.

Functional theories work as solvers for black-box problems (Loar, 1981; Schiffer, 1987). Given a system whose inputs and outputs are mediated by an epistemically inaccessible mechanism, the problem is to construct a theory which, given the input set, would be sufficiently explanatory and predictive of the output set. The solution suggested was to use the inputs and outputs as functional definiens for the "internal" (i.e., theoretical), states of the device.

Theoretically, the procedure follows the Ramsey-Lewis method for giving a functional definition of theoretical terms (Ramsey, 1929; Lewis, 1970, 1972). The advantage of the Ramsey-Lewis method is that it makes it possible to define a number of such terms together, through their relations to each other and to the non-theoretical vocabulary. In this way, the definition of these terms is given by reference to their theoretical roles in a given theoretical structure. A theory which gives us such functional relations we call a functional theory, and the result is a functional definition of terms, including psychological ones.

Applying this procedure to the case of the mental, this is supposed to work as follows. We choose first a psychological theory to supply the theoretical terms to be functionally defined. Here there are a few options: If you believe that FP *is* a theory, and that its generalizations have the status of laws (as I do), then you might also get the bonus of having defined our common sense psychological concepts for the same price. However, if you either think that FP is not a theory (Lewis, 1972), or that it is a false theory (Churchland, 1981, but of course, Churchland would resist any attempt to reductively define mental states), or that, in any case, cognitive science is our best shot at formulating an adequate theory for the explanation and prediction of behavior (Block, 1980), then you will use

it (cognitive science) as your source for functional roles.[39] In which case you will be called a *psycho-functionalist,* with the price to pay that it would not be our ordinary concepts you would be defining but some other scientific ones (not the concept WATER but WATER*). But be this as it may, whatever psychological theory one chooses, the technical procedure is the same.

Let us suppose that T is our psychological theory that specifies the causal relations obtaining between certain mental states, and between mental states and input and output states. We write it as a conjunctive sentence $T(s_1....s_n)$, with $s_1....s_n$ as *names* for mental states (as opposed to predicates. This requires some linguistic acrobatics). We then replace the names with variables closed under existential quantification to give the so called 'Ramsey sentence': $\exists x_1.....\exists x_n T(x_1.....x_n)$. This means that T, if true, has at least one realization.[40] A functional definition for each theoretical term can be then given by singling-out the position of this term in the Ramsey formula. In this way, the meaning of the theoretical terms can be determined with the only terms explicitly mentioned being those for the inputs and outputs. All the rest are mentioned "in absentia", as quantified over.

As implied above, some of the terminological choices of the theory depend on whether the theory is 'narrow' or 'wide'. For example, a Harman-type theory would include terms for worldly objects and affairs, and hence could consist in its empirical part of generalizations like the following: 'If agents, under normal conditions, are exposed to tigers, then they typically enter the perceptual state of 'I see a tiger in front of me'. The latter kind of state is a typical cause of the state 'I believe that there is a tiger in front of me', which is a typical cause of the state 'I desire to flee'. This, in turn, is a typical cause of the output state described as *fleeing.*
This theory can be then 'Ramsified' in the following way:

RS: ($\exists$x1,$\exists$x2,$\exists$x3,$\exists$x4)(x1 is typically caused by x2, which is typically caused by occurrences of tigers; x1 typically causes x3, which typically causes acts of fleeing).

In the case of a Block type theory, reference to proximal stimuli or outputs of sensory transducers would replace reference to tigers and other worldly objects, and reference to outputs of the motor cortex would replace reference to acts of

---

[39]By 'Cognitive Science' one should not necessarily understand any one theory or discipline but a broad range of scientific disciplines that attempt to give explanations of behavior.

[40]A modified Ramsey sentence can be also introduced in case we wanted to claim that T is *uniquely* realized: $\exists x_1.....\exists x_n T(x_1.....x_n)$& (y)(T(y)iff y=x).

feeling and the like. But be this as it may, we see that on both Block and Harman's functional theory, the idea of the mind as a causal nexus is the same, the differences are just in how *wide* this nexus is taken to be. And Fodor sums it up:

> Psychofunctionalism implies a model of the mind as a network of causal relations, where each node corresponds to a nomologically possible mental state and each path corresponds to a nomologically causal relation among the nodes... (1987, p. 77).

We can now go back to the question as to the justification for deriving the semantic properties of the attitudes from their functional roles. So far it was shown how to give a *functional* definition for mental states, that is, how those states ate to be defined by reference to their causal-role. But this is not as yet giving them a *functionalist* definition, one which registers in their semantic properties, and thus defines them *as* intentional states. In the case of our little Ramsified theory, I stipulated that some states of a system are to be given mental names, and then I went on to determine their theoretical role according to our functional theory. Indeed, this might work - if only in matters of methodology - if the system we describe had a finite number of states. But that is not the case with human systems which have an indefinite number of states.[41]

Well, as we saw from our discussion in chapter One, psychological theories give us more than just functional roles; they also supply a way to *constrain* functional role. That is, psychological theories work not only as empirical maps of actual and potential causal nexuses, but as a priori constraints on them, constraints which are expressed in the semantic relations between the propositional objects of the attitudes. And in effect, as we can now see, this is just the IT thesis I'm trying to define. We already saw Fodor's expression of this idea, so let me quote here another philosopher, Stalnaker, who says:

> to define a relation between a person or physical object and a proposition is to define a class of properties with the structure that makes it possible to pick one of the properties out of the class by specifying a proposition (1984:11).

It seems to me that what Stalnaker in effect says here is that the structural similarity between propositions and mental states, the propositional attitudes, is

---

[41]According to Block (1986), the number of thinkable sentences thirty words long is greater than the number of particles in the universe. It would be fruitless to try and construct a functional theory for all those states without some general constraints.

a kind of a *transcendental* condition on the construction of an adequate theory (the "class of properties") through which those definitions would be given. It is because of this condition that the content is made into an essential component in the definitions of the attitudes as psychological states and forms part of their structure.[42] Functionalism just takes this idea its own way by claiming that the said structural similarity is that expressed by IT, as it pertains to the structural relations between propositions and causal-roles. That is, we now see that it is because of the transcendental role of IT in the process of defining the attitudes that their functional definitions can be considered as also *functionalist* definitions. From this it appears that the black box of functional theories has in fact semantic colors, and thus the question now becomes, did IT and CRT solve between them the problem of intentional causation? I'm going to review now a few arguments against CRT, and will claim that some of them hit the mark more than others.

3. On Functionalism and Intentional Realism

The contention that the causal role of a mental state determines its content closes a circle for semantic functionalism. To wit, functionalism, a multiple-realization theory, came forward as a nonreductive physicalistic theory about the nature of the attitudes. The idea was to define "higher-level" properties, such as intentional properties, by reference to second-order physical properties such as causal-roles, and thus legitimize psychology as a science. If mental properties are defined as causal roles, then surely there is no question of their causal efficacy. Here two points present themselves for discussion. The first point concerns the question as to whether the solution offered by semantic functionalism is compatible with the basic tenet of intentional realism about the causal powers of semantic content. That is, we remember that the claim of intentional realism was not only that mental states are causally efficacious states, but that they are such *in virtue* of their content.[43] Simply put, the idea was that the semantic content of a mental state is taken to determine its causal role. But now we see that CRT's solution to the problem of intentional realism goes just in the other direction, by making a mental state's causal role determinative of its semantic content. And here it might be objected that since determination is an asymmetrical relation, then something must give. You cannot have it both that content determines causal role and that

[42]This, I would say, in a similar way to Davidson's constraints from charity as imposing a transcendental condition for the definition of linguistic meaning (in Davidson 1984).

[43]See footnote 1

causal role determines content.

Now, I think there is some truth in this objection in that some sort of asymmetric relation between the intentional and causal/physical realm is called for, as we indeed saw in the last chapter. However, from our discussion there we also saw that the issue about the relation between the mental and its physical realization goes beyond asymmetric *determination* relations, to asymmetric *dependency* relations, something which is overlooked, in my view by this objection. This point also connects with my general scepticism that determination relations are necessarily asymmetric for that depends, I would argue, on the kind of determination relations we are talking about. For example, suppose that we accept the reductive claim that the temperature of a gas is identical to the mean kinetic energy of its molecules. Then it seems to me that the gas being in a certain temperature *determines* its molecules having a certain mean kinetic energy as much as its mean kinetic energy determines its temperature. Now, one might think that there are differences between the senses of determination in both directions, the lower level *realizing* the upper level and thus *explaining* it. But the point is that this should not exclude the "higher-level" process from determining the "lower-level" one, even if that might be, and possibly is, a different kind of determination than through 'realization'. To conclude then, my point is that the asymmetry of explanatory and realization relations is consistent with the symmetry of certain determination relations, such as those invoked by identity theories, functionalism included. Because of this, identifying mental properties with functional properties is still consistent with the claim that content determines causal role, so long, that is, that we do not violate Hume's stricture. And this is the second point we need to consider.

Now, the problem vis-a-vis Hume's stricture that functionalism seems to face is that if content is defined in terms of causal role, then this makes the causal relations between mental states, under their intentional descriptions, definitional. If M is partly defined as the reliable cause of G, then 'M being a reliable cause of G' is logically entailed by 'being M'; and thus an instantiation of G would not be an independent event from that of M.[44] To make for the independence of G and M, one would have to show that they have other descriptions which are not logically connected. But in effect, this is something that is built into functionalism. What I have in mind is the idea of a multiple realization. Block says:

...a functionally individuated entity can, in principle, be identified by

---

[44]This is not the same situation as when 'being G' is entailed by 'being M' and the law 'M→G', since in such a case 'being M' and the law are themselves independent. In the case of functionalism, the law is entailed by the antecedent.

> independent (usually physicalistic) means and the mechanism of its causal connection to the effect described. For example, a gene identified functionally via the methods of Mendelian genetics can be identified as a clump of DNA via the methods of molecular genetics (1986:668).

Thus, to satisfy Hume's stricture, CRT depends on the availability of a realization level which is non-intentionally described and which bears no constitutive relations (either conceptual or through bridge-laws) to mental properties. But then again, this in itself would not be sufficient for the independent identification of the realization base. This is because, and as the discussion above makes the point, not just any physical state can count as a realizer of an intentional property. To satisfy the condition on being a (physical) realizer of an intentional state, a state has to show some further asymmetric dependence on the intentional in the way explained in chapter One. Thus, it should be noted that if we accept the basic idea underlying CRT that content is functional role (which I do), it cannot be sufficient, as it stands, for giving a complete definition for semantic content. That is, we see that it is not enough for a state to have the content it happens to have just in virtue of its causal relations, actual or possible, to other events, for the reason that not all of these causal relations can be on a par. What is required, in addition, is that some of those causal roles show a certain dependency on others. Thus, for example, we could say that although causing G is an essential property of being M, it cannot be a sufficient one. For we also require that in any world in which M is instantiated with its certain causal role (that of which 'causing G' is a proper part), that very causal role should also include 'causing G* (not identical to G)' which is dependent on its ability to cause G. But this means that not just any causal role of a mental states can constitute its content, only those that satisfy the asymmetric dependency condition. As we shall see in chapter Six, it is a virtue of the informational semantic theory, an atomistic version of CRT, that it does seem to satisfy that additional constraint from asymmetric dependence.

4. The Problem of Fusion

We saw that according to functionalism, a state's content becomes a function of its location in a causal network, or of the "node" it gets to occupy in it (see Fodor's graphical description on p. 43 above). The result is a system of states each of which corresponds to a unique mental predicate, the definiens, which function as these states' *names* (indices). The problem is though, that names lack structure, which would entail that mental properties would lack structure as well. For example, we remember that in our Ramsified theory above, I defined the predicate 'believing that there is a tiger in front of me' in terms of its causal dispositions to other states, which means that as a "node" in this causal nexus,

it is "fused" into the unstructured predicate 'believing-that-there-is-a-tiger-in-front-of-me'. And similarly 'desiring that there be a tiger in front of me' becomes 'desiring-that-there-be-a tiger-in-front-of-me' and so on. This consequence of functionalism is reminiscent of the fusion theory of mental predicates, as advanced by Quine (1960:214-5), according to which attitude verbs are treated as one syntactic unit with their content sentence.[45]

Now although this result is still compatible with the idea that the contents of mental states are propositions, because propositions lack structure, it would undermine the idea that the attitudes are relations to *mental representations*. For unlike propositions, mental representations are assumed to have a structure which then explains how they causally interact in ways that conform to that structure. Indeed, this is quite a common argument used by advocates of the representational theory of thought (RTT) to substantiate their claim in terms of the compositional nature of thought and mental processes (see Fodor 1987:135ff; Block, 1986).

Now to be sure, such arguments have a particular version of RTT in mind, which is the language of thought hypothesis (LOT). Other representational models, such as maps and pictures, lack such a compositional structure, since maps and pictures do not have proper parts. But if so, then the argument from fusion is just an argument against the compatibility of CRT with LOT.

5. The Argument from Holism, or Meaning Incomparability

Perhaps the most notorious problem for CRT is that of its purported holistic consequences. The argument goes something like this: Functional properties are those defined by reference to the causal dispositions of event tokens with respect to other events. That is, nothing can have a functional role unless it interacts with other events. The question now is, which events and which causal interactions? Well, according to Fodor (1987:64-5) and Fodor and LePore (1992), to distinguish those causal interactions of a mental state which are meaning constitutive from those which are not involves distinguishing those of its "epistemic liaisons" which are meaning constitutive from those which are not. And this seems just the distinction between the analytic and the synthetic which Quine has rejected (Quine 1953b:20-46). Conclusion, CRT implies meaning Holism.

Let's take this one step at a time, and first, a word about the relation between functional role and epistemic liaisons.

I said above that proponents of CRT wanted to define narrow content in terms of the dispositions of mental states to causally interact with other mental

[45]The term "fusion" is due to Dennett (1969).

states, and with stimuli inputs and behavioral outputs.[46] But I also noted that causal role is not yet epistemic liaisons, as the latter concern verification relations, not merely causal relations. Well, the point is that since CRT takes the narrow aspect of meaning to be non-representational, then given our transcendental constrains on the construction of functional theories, verification seemed like the most probable answer (see also Boghossian's suggestion on p.40 above to construe CRT's narrow meaning in terms of warrant). So this is how verificationism gets into the picture. But then this immediately invokes the classical dispute between Quine and the positivists.

To wit, not unlike proponents of CRT, the positivists claimed that a condition on the meaningfulness of a theoretical sentence is that there is a data sentence (the input or output specifications according to CRT) to which it is logically connected and by reference to which it is justified. Such data sentences were presumed to be the analytic implication of the theoretical sentences. In this way, the positivists tied down the meaning of a sentence to its conditions of confirmation and in so doing, advanced an *atomistic* conception of verificationism. Against this, Quine argued that "Our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (ibid., p. 41). That is, Quine has rejected atomistic verificationism in favor of holistic verificationism. Again, in terms of CRT, this means that if some of the epistemic liaisons of a mental state constitute its content, then they all do. The only way to block that slippery slope would be to introduce an analytic/synthetic distinction, that is a distinction between those confirmatory relations which are unrevisable in principle, and those which are not. But Quine argued that it is impossible to supply a principled distinction between analytic and synthetic sentences, ergo the holistic consequences that follow.[47]

---

[46]That was the narrow version. We also saw that there was the Harman, one factor, version which takes functional role out into the world. Other wide versions of functionalism might include Stich's fat syntax and Wilson's wide computationalism.

[47]I will later (p.114) have occasion to criticize Quine's 'conceptual' repudiation of the A/S distinction as incompatible with his thesis of confirmation holism, although that thesis itself is seen as a consequence of this very distinction. But one can argue for the no A/S distinction on a posteriori grounds, namely by reference to actual scientific practices. In which case we might still have to accept its holistic consequences to CRT.

To another point, Quine's identification of analyticity with un-revisability has been later attacked by Putnam and Kripke. If 'water is H20' is true, then it is

It was claimed by Devitt (e.g., 1996:117) that in the same way that we don't require a principled distinction between the essential properties of Mars (being a planet) and its accidental properties (having a certain atmosphere) for its individuation, then we should not require such a distinction to individuate contents. But this seems to me to beg the issue since it supposes that meanings are like planets, chairs, etc., that is, discrete entities. But why not suppose that meanings are rather like *being a number*, which is a holistic property?[48]

To conclude then, from the above it follows that CRT implies that the content of an intentional state is to be determined by reference to its confirmatory, inferential relations to *all* the other intentional states the agent concurrently has. But this in turn means that as far as its content is concerned, all those other states, on the confirmatory relations to them, would become its essential or constitutive properties[49]. Change one of them and you have thereby changed that state's content. Thus Block says:

> [I]f I say [believe] "Water is more greenish than bluish" and you say [believe] "Water is more bluish than greenish," then we have different narrow contents for "water" [WATER]...in the real world we can expect no two cases to be subsumed by the same law of content (Block 1991:40-1. Taken from Devitt 1993: 284).

Thus we see that functional role semantic theory implies content holism because

---

necessarily true and hence unrevisable in principle, even if it is not analytic in the traditional sense of the term.

[48] That number is a holistic property follows from F&L's definition of a holistic property. See below.

[49]This consequence is true in particular of mental-intentional states since, and as I implied in chapter One, it is not clear if phenomenal states have the kind of content that enters into epistemic liaisons.

Now, some would also question such sweeping holistic consequences on the ground of the modularity of the mental (Cf. Fodor 1983 were the issue of the modularity of certain mental functions is introduced). But we need to remember that the functionalist assumption, as discussed above, is that a theory has access to the causal-role of intentional states only by reference to certain semantic constraints about what *should* cause what. Otherwise put, functionalists see modularity as a possible consequence of functional theories, not as a constraint on them. Thus, if those semantic constraints show themselves to be holistic, as epistemic liaisons are, then I suspect that a functionalist will have to conclude against modularity.

it assumes that the inferential role of a mental state determines its content and because it is claimed that there to be no analytic/synthetic distinction. It follows, it seems, that all of a state's inferential properties become its essential or constitutive properties as far as its meaning goes. But since inferential relations are, indeed, relations, we get more than a holistic determination of content: we also get holistic *properties*. That is, we get a property such that 'if anything has it, then many other things have it' (Fodor and LePore 1992:2)[50]

Holism poses a threat to intentional realism since it implies what is often referred to as the problem of 'meaning-incomparability' (Cummins, 1996:30). The problem seems to be that if the having of beliefs about (e.g.) assassination depends on all other things the agent believes, then it would be very hard to find two people he could be said to believe anything in common about assassination (unless they were molecular Twins). The consequences of such a position to a nomic intentional science of the mind seem to be obvious. Because of that, not all proponents of CRT are willing to accept those assumptions of the theory which imply content holism.

For example, it has been claimed by Cummins (1996) that the argument for the holistic consequences of CRT rests on a confusion between inferential and functional roles. Cummins agrees that evaluating inferential relations should take into account what other beliefs the system has, accepting thereby Quine's confirmation holism. But, he claims, the determination of a functional role of a state does not similarly depend on the actual path the system instantiates, as opposed to its *possible* paths:

> What I can now infer from P, given what I now believe, might be thought to determine *P*'s epistemic significance for me now. But *P*s meaning is more plausibly identified with what I could infer from it given a variety of different cognitive contexts. In general, functional roles are always defined in a way that is independent of the state the system happens to be in, for the idea is to capture all the possible connections between states...not just the one the system happens to be in...When you write down a machine table, you have fixed all the functional roles, but you have not said anything about what state the machine is in (ibid.,pp. 34-5).

---

[50]F&L's idea of a holistic property is actually quite controversial since it ignores the point of context. Thus, according to Cummins, holism has to do with "a scheme of representations in which the meaning of each representation in the scheme is dependent on the meanings of all the others" (1996:35 n9). This leaves it open that a property which is holistic in one context or scheme would be atomistic in another.

Hence, Cummins continues, if CRT identifies meaning with functional role rather than with inferential role, then holism should not follow. Indeed, Cummins maintains that making the content of MRs derivative on the inferential roles of the attitudes is not even possible since the two in fact cross- classify each other. To make his case, Cummins makes a distinction between being a propositional attitude with an application content, and being a mental representation with a representational content (ibid., p. 16). An application content is the content of a state with some representational function (an 'intender') on a particular application of a representation. For example, if the function of the intender is to represent the current temperature, and the representation it uses is '5 degrees Celsius', then the application content is 'the current temperature is 5 degrees Celsius'. But then this allows us to see that the content of the representation is only a component in the application content, and thus in fact independent of it.

Cummins claims that if the content of mental representations is independent of the content of the attitudes which make use of them, then inferential role cannot be constitutive of meaning. The idea is that inferential role is in the first instance a property of propositional attitudes because it is only they which are *epistemically* assessable, whereas MRs are *semantically* assessable. The reason is that the attitudes enter into epistemic relations only as attitudes of a kind: For example, your belief that you are eating a cake could be justified by a certain perceptual input that involves your eating a cake, but then this same input would not be one from which your desire for a cake will be derived. Rather, we expect that this input will result in the cessation of that desire.[51] Cummins relies on considerations such as these to show that inferential role cannot determine representational content because MRs can be the objects of different kinds of propositional attitudes. But if inferential role does not determine MRs' content, then the constraint from simultaneity evaporates and with it the problem of meaning incomparability. To wit, the latter problem can arise if what one believes depends on everything else they concurrently believe. Whereas, it now seems that what one believes, that is, the content of the MR which is *used* by this attitude, depends only on the systematic contribution of that MR to the inferential role of the attitude. And as this systematic contribution includes possible as well as actual instances, it does not succumb to the problem of

[51]Indeed, it would seem that by making the distinction between application content and representation content, Cummins merely distinguishes between an inferential role conception of content and a truth-conditional one. For as we can see from the cake example, although the justification conditions (what I called their 'epistemic liaisons') on the belief and the desire would presumably be distinct, their *satisfaction* conditions are the same.

meaning incomparability.

Cummins' argument should be seen as anti-Quinean in two respects. First, by dissociating the conditions on content determination from inferential or confirmation relations, Cummins goes against Quine's verificationism (though not against Quine's confirmation holism). Second, by distinguishing the functional role of MRs from the inferential role of the attitudes, Cummins seems to make the point that revisability and unrevisability are not just epistemic issues. For we now have a notion of content, and hence of inferential relations, which is independent of the current epistemic economy of the agent. Of course, epistemically, one can "hold onto any statements, come what may, in the face of recalcitrant data" given one's other beliefs and desires; for example, the desire to hold onto a statement, come what may, in the face of recalcitrant data. But one can argue that this has nothing to do with *actual* confirmatory relations which are relations between concepts, not between beliefs. For example, one can believe that the rusting (corrosion) of a metal in air confirms the liberation of phlogiston, whereas what it actually confirms is the oxidation of the metal. Thus, real confirmation, as opposed to putative confirmation, has to do with connections between concepts rather than between the propositional attitudes incorporating those concepts. I might call this notion of non-epistemic confirmation "wide confirmation", and the theory that invokes it, "wide verificationism".

The reason why all that is important to our discussion is twofold. First, if the inferential role of the attitudes plays no role in content determination, that is, if only functional role does, then CRT loses our interest as a theory which can substantiate IT. For we remember that CRT seemed to be our best shot at substantiating the relation of isomorphism between the functional role of the attitudes and their inferential role, and that due to the assumption that the inferential role of a state plays a constitutive role in its determination of content by functional role (see the transcendental condition on p. 44 above). This also brings us to the second reason why this discussion of wide verificationism is important to us.

The point is that once we lose our transcendental condition for the determination of content by reference to functional role, we are in effect left without a reason to justify such a reduction of content. That is, we are now back to the question as to what can make functional role a determinant of content. For surely it is not enough just to say that content is functional role, in the same way that it is not enough just to claim that D defines C. One also has to come up with a justification for it. In the case of classical, or (narrow) CRT, the justification came in the form of the transcendental condition on the construction of functional theories. But once we lost that, we lost our justification for reducing content to functional role.

But now, once Cummins allows a new kind of verificationism to replace

the old one, wide verificationism, we might find here a way to regain our transcendental condition, albeit in a somewhat different form. As said, the idea behind wide verificationism is to determine content not by reference to epistemic liaisons but by reference to the confirmation relations between the concepts themselves, independent of whatever attitude they find themselves in. This would still mean that when we go about constructing functional theories we should constrain ourselves by inferential relations, but this time, we will have to find a way to distinguish between those inferential relations which are accrued to a system from the concepts it uses, and those which are a result of the contribution of the latter "mixed" with those of the contribution of the epistemic liaisons of the attitudes themselves. That is, we need to find a way to distinguish those inferential roles which are associated with representation content, and those which are associated with application content, in Cummins' terminology. To see how one should disentangle the two kinds of inferential relations, and also the implications of this procedure for a theory of mental content, I turn to the next section.

6. Ideal Verificationism (or 'Idealism')

According to Cummins' theory of representational meaning, the meaning of MRs is a function of their systematic contribution to the inferential liaisons of the attitudes, actual or possible. That is, we now look at the meaning of a mental representation, or the concept it expresses, by reference to more than just actual psychological systems in which a concept plays its functional role. Rather, we look at possible psychological systems as well, those in which that concept *could* play such a role. This certainly would be able to carve a distinction between the contribution of a concept itself to inferential role and its contribution given the particular epistemic organization of the system into which it is incorporated. However, leaving the door open for an indefinite array of unconstrained possibilities in which a concept could play a contributive role, as Cummins seems to be doing, will preempt the project of defining these concepts in my opinion. For surely, any one actual psychological system, or even the totality of those in the actual world, constitute only a negligible fraction of the possible totality, which would make the transcendental condition empty of content. For surely there are now no longer any meaningful constraints as to how to decipher the content of the black box, except, of course, those relating to the actual world and those close to it.

However, there should be in fact no reason to adopt such an extreme holism since it can be easily acknowledged that many, if not most, of these possible psychological systems are quite irrelevant to the independent contribution of concepts to inferential role. For surely, the less truthful a system

of beliefs and desires is, that is, the less it tends to represent the world (including the organisms true needs) as it is, the less the independent contribution of its concepts can be felt. Because of this, what we should be looking for are not really all possible psychological systems in which mental concepts can play a role, but only that subset of which where psychological systems operate at their full potential, that is, at their *ideal.*

The idea is this. We are trying, with Cummins, to get away from functional theories which attempt to define mental content by reference to the concurrent intentional states of an agent, those which form a rational cognitive economy. Instead, we are seeking to map the systematic contribution of the concepts which, ex hypothesi, determine the inferential role of the attitudes, and that by reference to possible theories as well as actual ones. But I implied that it would be otiose to refer in that way to *all* possible cognitive economies including those which include extremely bizarre and outlandish intentional states, since these could tell us very little about the concepts that *we* use. If anything, it would distort the independent contribution of mental concepts rather than expose them. Instead, it would make more sense to attend to those cognitive systems in which the systematic contribution of the concepts themselves is as close to that of the corresponding properties as possible. But this would just mean reference to those cognitive systems which incorporate the true theories and beliefs, including such of their external extensions as the use of reliable instruments and experts, for the confirmation of one's concepts (cf. also Rey 1993 who also suggests ideal verificationism). This is, after all, what I would understand by a psychological system which operates ideally.[52] That is, we can now say that instead of focusing on the actual operational conditions of psychological systems, we focus instead on those conditions under which those systems (such as the human psychological system) would operate to their *full extent.*

Let us assume that it might be possible to interpret Cummins's theory of 'representational content' in this "idealised" way.[53] The question then becomes,

---

[52]One should not confuse this notion of ideal conditions with that advanced by certain teleological and teleo-functional theories. For example, I make no claim that ideal conditions for the operation of a psychological system are also ideal from an evolutionary point of view. For as is well known, sometimes believing falsehoods has a better survival value than believing truths. Still, one would not say that in this case the psychological (cognitive) system was operating at its full potential.

[53]In fact, it does not matter anymore if there is any interpretational relation between ideal verificationism and Cummins' theory. At this point I use Cummins' theory merely as a heuristic device.

whether the conceptual content we get via such an idealized functionalist theory is really representational (i.e., truth-conditional) content, as Cummins seems to think? Well, when I said above that ideal verificationism gives *true* confirmation of concepts, I surely did not mean that this truth is semantic truth. The reason for that is that even in the case of 'ideal' confirmation, that is, the case of inferential relations between concepts which supposedly obtain independently of the actual theory (or mental economy) which makes use of them, reference to *some* theory is still presupposed. That is, since even ideal theories are just that, theories, and since even ideal confirmation is just a kind of *confirmation,* whatever relations are found to obtain between concepts in ideal theories still falls short of the relations between the worldly properties themselves.[54] Because of this, it is still left open that some relations and distinctions in reality are not confirmable even by an ideal theory, because what is ideal for human intentional systems might still not be the *real* (I assume that this is the reason why we draw a distinction between the two concepts). Another way to put this point is by saying that the ideal *for us* is what we can know when we have exceeded all our physical and psychological contingencies, which is to say, when we have used our intellectual capacities to their nomic potential. It does not mean that in ideal circumstances we would come to know what God, or any omniscient being, knows. The latter is not an ideal but a fantasy.

Drawing the distinction between the real and the ideal enables us to conclude that the notion of truth we use in wide verificationism is still narrower than that of semantic or extensional truth, in which the only thing which matters to truth is that to which the concept applies, independently of *any* actual or possible theory in which that concept is used. On the other hand, when we recede away from this extreme (and in my view implausible) realism, we can have a more benign form of verificationism which differs in scop from that of classical verificationism. But I will leave the more detailed discussion of the sense in which concepts are essentially theory oriented, and of the way their identity conditions are limited to the role they play in nomologically possible theories, to chapter Six.

To summarize then, I have suggested that we can extract from Cummins (1996) the theory whereby we can define the concepts expressed by mental representations by reference to ideal but still nomologically possible verification relations. I have called this 'wide verificationism' because it makes the inferential relations between concepts independent of the particular psychological system

---

[54]This shows, in my view, the fallacy in those views which take concepts to be reducible to properties. The mistake lies in that concepts, unlike properties, are essentially theory oriented. I will discuss one such position, that of Bealer (1982), in chapter 4.

into which they are incorporated. On the other hand, it is not completely 'wide content' either since I have claimed that all conceptual relations are dependent on their incorporation into *some* theory or other, or *some* psychological economy or other. This will mean that the kind of content that this theory would give us is not extensional or truth-conditional content, and hence still narrow to an extent.[55]

Now, if we could adopt wide verificationism, or idealism, as our new format for CRT, we could certainly regain our hold on IT, and even be at a somewhat better position than we were in with respect to classical (narrow) verificationism. This is because we could now establish the isomorphism relation directly between concepts and their functional role, independently of the particular role that the attitudes play in actual mental economies. As said above, in this way we could also avoid to a significant extent the problem of meaning comparability, although the definition of concepts are still holistic: They depend on the *totality* of inferential relations that those concepts licence in ideal theories. The problem with meaning incomparability does not follow since, given a more or less of an identity amongst humans in their psychological possibilities for ideal functioning, there will be an 'across the board' identity in the functional role of the concepts they use. On the other hand, there might be, and probably is, a problem of meaning incomparability with creatures of *other kinds*, but that is not something that should strike us as particularly surprising.

From the above discussion, it certainly seems that I find favor with many features of Idealism as opposed to classical CRT. It is time, though, to show that there are also some serious problems with this theory, some of which I will mention now, others, which also concern more generally classical CRT, I will discuss in the next section.

The first problem I have with Idealism concerns the possibility to express this notion of content. For even if one does not think that the content of our concepts should be transparent to us, in the sense that we would be able to point to the inferences those concepts licence, we still want to think that we can in principle understand them. For example, we want to think that although people

---

[55]I cannot resist the feeling that many of the difficulties in understanding Davidson's theory of radical interpretation arise due to the fact that Davidson ambiguates between two notions of meaning. On the one hand he advances a truth-conditional notion of meaning which he associates with Tarski's semantic conception of truth, and on the other hand he gives a theory-relative notion of meaning, relativised to a choice of an interpretational theory for a language. But once a theory is 'widened' in the sense that it does not depend on particular applications of concepts (meanings), the gap between the two notions of meaning begins to narrow down.

in 1750 didn't know the full content of WATER, that is something that could have been remedied. Just discover its scientific essence and spread the word ("it is H2O"). But now it seems that there is just no way to ever being able to tell what WATER means. How could you, given that it means an indefinite array of theories and other verification procedures which are the possible or *ideal* ones? Thus, this suggestion just amounts to the claim that content is inexpressible.[56]

Then there is this issue. Suppose that WATER means all the ideal verification methods that result in confirming that something is water. Then, surely, many of these are redundant. For example, think about all the actual and possible experts in the world who can assist one in that verification. All these experts should count as means of verification, and hence should all be part of our content of the WATER concept. Or if you think that all experts are one method of verification (even if they might have different ways of detecting water?), then surely there is an indefinite number of possible ways to tell that something is water which are not of the same kind (unless they all count as 'methods for detecting water'). In chapter Six I will claim that there is something on the right track in Idealism, namely that what we mean by a concept has something to do with all the possible ways to track its instances. But I will claim that the *methods* we apply, actual or possible, should be quantified over. Indeed, the point will be that what we should care about concerning the content of our concepts is not so much how we verify their application, but what it is they apply to, given our psychological potential for doing so.

7. The Argument from the Underdetermination of Content

The last two objections that I posed for CRT concerned a specific version of it, which I called Idealism. In this section I'm going to review two arguments that bear on any form of CRT, to the effect that using IT to constitute mental content would underdetermine it. First I shall view an argument according to which the logical form of laws is too "coarse"[57] to be used in the determination of content.

---

[56]Compare to Fodor's contextual notion of narrow content which, by Fodor's own admission, cannot be completely expressed in any one language (in Fodor 1987:50).

[57]The metaphorical terminology of "coarseness" or "fine-graininess" can be best explicated via the phenomena of Intensionality and Extensionality in logic and language. Thus according to Bealer, intensional entities are "..entities that can be different from one another even though they are identical in extension" (1982, p. 30). Another way of saying this is that intensional entities are always more "fine grained" than their extension. Thus I take it that for any

That is, this part of the argument challenges the idea that functional role, as expressed in law sentences, can have enough structure to substantiate IT. Then I shall review another argument according to which it is inferential relations which are too "coarse" to determine causal relations between mental states.

A. The Problem of the Generality of Laws

We saw above that the strategy of CRT is to find a confirmed empirical theory from which the functional role of mental states could be extracted. Since in such theories the causal order, as expressed in the law sentences of the theory, is assumed to be aligned with the semantic order, the reduction of content to causal role is assumed to be accomplished. The problem is, though, that psychological laws, which express the causal relations between psychological states, are apparently too general for the purpose designed for them here (Fodor, 1987:70). That is because such laws merely *quantify over* the content of mental states, which means that they do not make any specific mention of them. We can see this when we attend to our example of the (purported) psychological law from chapter One:

PL: '*ceteris paribus*, people who desire *p* and believe that *not-p unless q*, tend to behave in a *q* manner'.

Surely, the law PL expresses a partial causal role of the attitude subsumed by it, but it does so for *any* belief and desire with the same logical form, irrespective of what *p* or *q* mean. That is, PL would express the same functional role if *p* is the proposition that snow is white, grass is green, or whatever. But if that is the case, then a functional theory which relies on such crude law sentences could not become a functionalist theory, that is, it would fail to determine contents at the level of "grain" assumed for them under the FP scheme.

It seems to me that this problem should work also against Cummins' idea that the content of MRs can be given by reference to their functional role in the cognitive economy of mental systems. For unless Cummins would be able to supply a procedure to determine functional role independently of psychological laws, that is, under an individuation scheme which is more "fine-grained" than that of laws, his theory faces the same problems. And of course, the same would be true for idealism which also makes reference to laws, although ideal ones. Because of this, we can now see that avoiding the problem of meaning incomparability is not the only thing that should worry a functionalist.

---

two entities (or expressions) a and b, to say that a is more "fine-grained" than b is to say a stands to b like an intension to its extension.

It might be suggested that the way for the functionalist to go is to find some method of individuating content before it enters the scope of the intentional operators. For example, if reference to propositions were available in the input or output specification, then it might be possible to *quantify-outside* of the intentional context as follows:

(P): Let *p* be the proposition that there is a tiger in front of *S*. Then, if there is a tiger in front of *S*, then *ceteris paribus*, *S* comes to believe that *p*.[58]

The problems with this suggestion, though, is that it is no longer "narrow content" functionalism, nor even the sort of "wide content" functionalism I discussed in the previous section, since reference to worldly affairs and actions extensionally construed is introduced. It also means that if CRT adopts it, then the narrow component of content is no longer distinct from the referential component. In which case it is hard to see what reason would remain for us to adopt CRT rather than just a referential theory of content.

B. The Problem of Equivalent Propositional Attitudes

Another "grain" problem arises from the issue of equivalent attitudes. Equivalent propositional attitudes are those attitudes which satisfy two conditions: They belong to the same attitude kind, and they are directed to equivalent propositions, that is, propositions which have the same inferential role. An example of equivalent propositions would be those expressed by the schema ~(pvq) and (~P&~Q).[59]

The problem with equivalent propositional attitudes arises when we attempt to project their causal roles by reference to their inferential role, as follows from IT, and capitalized on by CRT. If inferential role is assumed to determine causal role, then IT should determine psychological natural kinds. For example, we would expect that the propositional attitudes individuated by reference to ~(Pvq) and (~P&~Q) would have the same causal role, since ~(Pvq) and (~P&~Q) have the same conceptual role: They enter into the same patterns of inferences. The problem is, though, that this seems to fly in the face of

[58]For a discussion of such idea see Schiffer, 1989:33.

[59]When they are relativised to the same logical contexts. Thus, for example, if (~p&~q) has in its logical context (~p&~Q)⊃R as a theorem, its inferential role would then differ from that of ~(pvq). Thus, again, we are assuming an *idealized* context.

empirical data: people do not infer all the consequences of their thoughts, even if these are opaquely construed. That is, empirical evidence suggests that equivalent propositional attitudes might nevertheless differ in the role they play in reasoning, decision making, and other psychological processes, *even when the contribution of other attitudes is nullified* (that is, even under ideal conditions). To paraphrase a remark by Fodor, it seems that the attitudes are not closed under the consequence relation. This means that, for example, we could not predict that if x believes that ~(Pvq), and ~(Pvq) iff (~P&~Q), then x would come to believe that (~P&~Q).[60]

Now to be sure, and as follows from the idealization clause, the problem with equivalent propositional attitudes is in effect a problem with equivalent *concepts*. The problem seems to be that the inferential relations between concepts are too "coarsely grained" as far as the empirical adequacy of CRT goes. In other words, it now seems that inferential roles, even idealized ones, are too "coarse" for the individuation of the causal-roles of propositional attitudes. But this appears to defeat the whole purpose behind CRT, which relies on the isomorphism between the two domains to substantiate the content of those very concepts. To repeat, the idea behind CRT was that the inferential role of the attitudes constrains their causal-role such that equivalent propositional attitudes form a psychological kind. But now we see that it is not the case that equivalent propositional attitudes have the same causal role, and hence that inferential role cannot classify equivalent propositional attitudes as a psychological kind.

From this discussion it would seem that a more discriminative scheme for the attitudes is called for, one which would be able to predict the differences in causal-role which the data suggests. In other words, we seem to need a notion of content that would be empirically adequate, and that seems to point beyond inferential role of either type.

8. Content Is Not Inferential Role

Overall in this chapter I have reviewed a number of difficulties facing CRT, all arising from the main dictum of CRT that content *is* inferential role, narrow or "wide". I think it is now time to present an alternative to this theory, but one which still respects the idea that there are constitutive relations between content

---

[60]This issue should be distinguished from those cases where people fail to *form an attitude* towards a proposition which is closed under the relevant attitude relation. For example, an agent can believe that ~(Pvq) and believe that ~(Pvq) iff (~P&~Q), but fails to form the belief that (~P&~Q). I think we should see these latter cases as indeed exceptions to how people usually make inferences and thus as showing irrationality.

and inferential role. Well, the alternative to content being inferential role can be that content *determines* inferential role, whereas inferential role does not determine content. Instead, content is determined by the worldly properties themselves. As we shall see in the following chapters, this solution is better equipped to solve most, if not all, of the problems that CRT faces. Let me elaborate on this a bit.

So far in this thesis I have accepted as a basic starting point the claim of intentional realism that mental states have their causal role in virtue of their content. What in effect this statement means is that having the same content implies having the same causal role. Or alternatively, that a difference in causal role implies a difference in content. But then we saw in this chapter that according to CRT, content is determined by inferential role, which is in turn identified with functional role. In this way, content is reduced to functional role on the assumptions, taken by CRT, that 1) content is inferential role and 2)inferential role can be reduced to functional role. Thus it follows from CRT that a difference in causal role implies a difference in inferential role, or alternatively, that identity in inferential role implies identity in causal role. But notice that from the perspective of intentional realism, it is not necessary that identity in inferential role would imply identity in content. For, if inferential role did not determine content, then two concepts might have the same inferential roles, but still be different in their content. And then, it would be also possible that two concepts have the same inferential roles but distinct *functional* roles, because they differ in their *content*, content being that which determines causal role directly ("over the head" of inferential role, so to speak). In which case, it would be possible to explain how equivalent attitudes can be identical in their inferential relations (by definition) but exhibit distinct causal roles, as the data shows. That reason being that they differ in their *contents.*

My suggestion is therefore, that we adopt a conception of the individuation of the attitudes for the purpose of psychological explanations and predictions that distinguishes states not by reference to their inferential roles, but by reference to their content, whatever that turns out to be, *given that these states differ in their causal role.* That is, my suggestion is to start from "below" in identifying those attitudes which are different in their contents by reference to their distinct causal roles (of course, not an easy task since we need to do quite a lot of idealizing), and then determine their content independently. This is a kind of inference to the best explanation, *or a transcendental deduction of the attitudes*, but which will be substantiated through an informational theory of content.

Here are, therefore, the lessons I think we can take from the discussion of IT in this chapter. We saw that the idea of substantiating IT through the reduction of content to causal roles had some unacceptable implications. In particular, those relating to the problems of CRT concerning meaning

incomparability and the problems with the "grain" of content. That last problem also bears on the empirical adequacy of a theory of content based on inferential role. My claim was that the kind of content we should be looking for needs be atomistic, that it needs to make reference to laws which connect intentional states with properties in the world, and that it needs to be with a level of "grain" that the data supports. We shall see that all these desiderata are satisfied by an informational, truth-conditional, notion of content. The place to begin the search for this notion of content is, I believe, with the semantic theories of Frege and Russell.

Chapter Three

The Theory of Cognitive Values

1. Some Presemantic Considerations

My discussion in the last chapter of the isomorphism thesis (IT) presented it as a descriptive thesis about the relation between the semantic properties of the attitudes and their causal roles. The assumption was that as a matter of fact, the causal roles of the attitudes mirror their semantic relations, by which we had in mind conceptual or even broader inferential relations. As the discussion progressed, it was clear that such semantic relations were traditionally construed narrowly (without reference to worldly objects and states of affairs) rather than widely (with such a reference), something which made it all the more natural for conceptual role semantic theory (CRT) to get into the picture. For CRT (the 'short-arm' version which stops at the skin) is just the attempt to *define* the narrow inferential properties of the attitudes by reference to their internal causal role, and thus solve the problem of isomorphism *literally* by definition.

However, we saw that taking this route burdened IT with some of the more notorious problems that CRT has been found to face, mainly those concerning its holistic makeup. Such were the problems from meaning incomparability and fusion, the latter resulting in its incompatibility with RTT.

However, and without taking lightly any of these problems, I think that the more pressing problem for CRT to our discussion lies in its failure to adapt to a conception of IT which takes its cue from the normative aspect of the Folk PEPs, rather than its descriptive aspect. I will explain.

One of the parameters by which Folk psychology functions in assessing and explaining other people's behavior is by producing *standards* for correct behavior and thought, not just by making educated guesses as to what might or might not happen. That is, we expect other people to conform to certain canonical forms of logical thinking, including the formation of certain beliefs under standard conditions, and the production of an array of behavioral outputs, given people's interests and the state of the world at that time. Otherwise put, we expect people to be rational in the broad sense of the word which includes their long term interests (and those of their peers), and given what the world, in which they function and survive, is like .

Well, one of the facts of life which the Folk recognize is that people cannot be asked to know everything, even of those facts which are quite relevant to the satisfaction of their needs and wants. For example, the Folk recognize that

although people can truly believe the sentence
(1) Leningrad is pretty,
they might fail to believe the sentence
(2) Petrograd is pretty,
even given the truth of
(3) Petrograd is Leningrad.
That is because they might just fail to believe (3), and thus fail to make the right connection between (1) and (2).[61] We can imagine that such a person would have two mental "files", one under the name of "Leningrad", the other under name of "Petrograd", each containing a stock of information about the Russian city. But because of a failure to identify the two files, it would seem that the 'cognitive economy' of this individual would be in some disarray, with some obvious costs.

Recognizing the utility in the proper classification of the information cognitively available to the subject, the Folk expect people to be able to follow more than just narrow inferential connections but wide ones as well. For example, the Folk expect individuals to infer (2) from (1) given the truth of (3) just because they think that the logical or cognitive economy of the mental would increase as a result of that. Surely, coming to learn about the truth of (3) would extend the epistemic capability of agents as it would allow them to hit upon a new stock of true beliefs at one stroke - all those beliefs formed from crossing the information in one file with that in the other.

Following such wide (i.e. extensional) semantic constraints would enable the agent to avoid not just cognitive or logical inconsistencies, that is, inconsistencies which in part depend on how the information is *represented* to the agent, but informational inconsistency as well: one which has to do with how things connect in the world. For example, we might suppose that while the sentences
(1) Leningrad is pretty,
and
(4) Leningrad is not pretty,
are both cognitively and informationally inconsistent, sentences (1) and
(5) Petrograd is not pretty,

---

[61]Or alternatively, they might believe (3) but fail to connect it to (1) or (2). But the latter possibility is a concern, if at all, for what I call below 'narrow rationality'. If at all, since it can be claimed, as I did in chapter Two, that a failure to infer (2) from (1) given a belief in (3) shows irrationality, and hence should be treated as an exception. Or more generally, I say that if an agent believes a proposition Σ, and believes that Σ entails Ω, but fails to believe Ω, then he or she is irrational and should be considered, from the point of view of psychology, as an exception.

are "only" informationally inconsistent. On some of our norms of rationality, people who hold both (1) and (5) (but not (3), the identity sentence) would not be "censured", but people who hold both (1) and (4) will be. We might call this 'narrow rationality' or alternatively, narrow constraints on rationality. On other norms of rationality, those which concern not just cognitive consistency and coherence but also truth, hence 'wide rationality', people who hold (1) and (5) would also be "censured". The reason for this added grain of rationality is that although the Folk are sensitive to the "internal" considerations which narrow rationality governs, they also expect agents to be sensitive to the informational aspect of their mental states.[62] This is shown by the fact that it is common for the Folk to explain and predict behavior by reference to widely recognized names of individuals and kinds rather than their phenomenal properties or other purely conceptual descriptions. Put another way, it would seem that the Folk recognize that logical or cognitive consistency is not sufficient for being rational. One has also to be in possession of the truth.[63]

Now in chapter Two we saw that some versions of CRT have divided the meaning of an expression, and the corresponding attitude, into a narrow aspect, determined by inferential role, and a wide aspect, determined by a truth-conditional referential component. This was the two-factor theory of Block, McGinn, Field, and others. In this way, two-factor theorists have attempted to produce a viable theory of mental content which could both satisfy the constraints associated with Fregean semantics, that is, that it is possible for individuals to be rationally related to distinct and incompatible attitudes towards the same object, and those imposed by Putnam's Twin-earth thought experiments to the effect that meanings 'in the head' do not determine reference. However, the bifurcation of content into two independent factors has its own problems to deal with. In particular, it now faces the challenge of how to align

---

[62]We can imagine the reaction of a group of English scientists when they come to believe that one of their colleagues thinks that water is wet but H2O is not. After making sure that they have interpreted him correctly and that he has no new theory about water (or about constitution not being identity), they would just conclude that he lost his mind.

[63]The claim that Folk psychology is normative and that it makes reference to wide constraints of rationality can be found in Davidson (1970, 1984b), Dennett (1987a,) Stich (1978, 1983), and Burge (1979, 1986). Scholars who accept the normative aspect but claim that FP is narrow rather than wide include Searle (1983) and Loar (1988). In any case, we should distinguish the claim that FP uses wide norms of rationality from the claim that the individuation scheme of the attitudes is either constitutively normative, or wide, or both.

the two aspects of content with each other (Fodor and LePore, 1992; LePore and Loewer 1987). As F&L say, "what prevents there being an expression that has an inferential role appropriate to the content *4 is a prime* but the truth-conditions appropriate to the content *water is wet?*' (Ibid., p. 170). And McDowell has complained that the Cartesian picture of two independent domains would immerse the interior with subjective darkness:

> How can we be expected to acknowledge that our subjective way of being in the world is properly captured by this picture, when it portrays the domain of our subjectivity - our cognitive world - in such a way that, considered from its own point of view, that world has to be conceived as letting in no light from outside? (1986:160).

Now, Block, for one, thinks that there is some way for inferential role (or narrow content) to determine reference in a manner which is compatible with Putnam's constraint, and that is to determine more generally the *kind* of theory of reference for the expressions of the language (and similarly for mentalese). After all, to take one notable example, Kripke's considerations for the direct theory of reference for proper names are themselves expressed purely conceptually and are, therefore, "narrow" in the sense defined here. After all, one does not come out of reading "naming and necessity" knowing to what "Socrates" or "Moses" refer to, nor is such knowledge required in order to understand Kripke's arguments against the descriptive theory. In the same way, one could rely on general conceptual considerations to conclude that proper names (e.g.), are directly referential without at the same time specifying what object each and every names denotes.

The problem with this suggestion is, as F&L and others noted, that it does not issue in a theory of reference, as Block would like us to think, but in just another theory of meaning. Indeed, I think it is illuminating to compare Block's mistake to that of taking, as many do, Kripke's own sociohistorical theory as a theory of reference. As Almog insists (1984:482), what is determined in sociohistorical chains are not the semantic facts about an expression, that is, about its reference, but instead *a way of assigning reference* - something that can be seen as part of the *linguistic meaning* of the expression. Because of such considerations, Almog takes historical chains of linguistic use to constitute only "presemantic" facts about a language, as opposed to determining the semantic properties of expressions such as reference and truth. My claim is that the same is true for Block's theory, that narrow inferential role can determine at most the kind of theory of reference for a language or mental states, although the inferential role Block invokes lacks Kripke's historical aspect. But if Block's theory is only "presemantic" as we now seem to see, then it cannot supply the representational or wide content that IT requires for it to conform to the Folk

explanatory scheme.

At this point we seem to be left with two options for a theory of content which is compatible with the Folk normative psychological scheme. One option is to try to revive the Fregean alternative which, according to the received wisdom, advances a notion of content that is both "enriched with Fregean fineness of grain" (McDowell, ibid.) and at the same time representational. But then the question would be how to rescue it from the Putnamian objection that what's in the head cannot determine reference. The other option is to "go wide", in the form of some truth-conditional semantics for thoughts but which does not violate the narrow constraints on rationality. One theory that seems to satisfy this model is Russellian semantics, but as I will show, at a price of losing the world as we know it. However, I will claim that there is a *third* alternative which combines aspects from both Fregean and Russellian semantic theories. The idea would be to construct a semantic theory which would replace the Russellian logically proper names with Fregean names, given the appropriate semantic restrictions, with the result that the differential references of such names, as in Russellian semantics, would become their cognitive values. This will give rise to a conception of truth-conditions for thoughts which are as fine-grained as are Fregean senses. To see how that would work, we need to plunge into the deep seas of both semantic theories.[64]

2. Frege's Theory of Meaning[65]

According to the accepted wisdom, Frege's theory of meaning arises out of

---

[64]To anticipate, my theory is not quite that of McDowell (1977 and ibid.) who also suggests a theory that combines aspects from both Fregean and Russellian semantics. In particular, McDowell wants to advance a notion of thought based on his interpretation of Fregean senses as object dependent. However, eventually cognitive differences in McDowell's theory hinge on differences in *truths*, not in *truth-conditions*.

[65]The question as to the kind of theory of meaning which Frege has advanced in his post 1892 writings is perhaps the most contentious of all in the analytic philosophy circles. Although in what follows I will make some exegetical remarks, I ask the reader to take them in the dialectical spirit in which they were written. In particular, by following Salmon's interpretation of 'Frege's puzzle' as 1) identity transcendent and 2) involving informational differences, I take a somewhat liberal approach to Frege's own discussion. Indeed, to do justice to what was Frege's own view about the matter would involve taking much more of his corpus into account than I can hope to achieve here.

considerations which are sometimes known as 'Frege's puzzle'. The puzzle arises once one assumes two apparently innocent semantic principles. The first, which is usually associated with Mill's theory of names (also known as the 'Fido'-Fido theory), is that the meaning of a name is its referent. The second is the principle of compositionality, according to which the meaning of a complex expression, such as a sentence, is a function of the meanings of its constituents together with their modes of composition. Given these assumptions, it seems impossible for two true identity sentences such as "a=a" and "a=b" to differ in their meaning, or what Frege called their cognitive value (Erkenntniswerte). To use an example with undisputable instances of proper names, the question is: how can the sentence "Hesperus is Phosphorous", if true, differ in its cognitive value from the sentence "Hesperus is Hesperus"? By the "Fido"-Fido theory, "Hesperus" and "Phosphorous" have the same meaning, and since the two sentences are composed in the same way from their parts (i.e., their syntax is the same), their meaning should also be the same. But according to Frege, that "a=a" and "a=b" differ in their cognitive value is *shown* in that the first sentence is known a priori and is analytic, while the second "often contains very valuable extensions of our knowledge" (Frege 1892b:56).[66]

In effect, Frege offers here 3 distinct criteria for identity or difference in cognitive value, the first presupposing a distinction between a priori and a posteriori sentences, the second concerning the analytic/synthetic distinction, and the third in terms of 'valuable extension of our knowledge'. However, given that I have already expressed my sympathy with Quine's thesis of confirmation holism and his refutation of the analytic/synthetic distinction, the first two criteria cannot be an open option for us. One might attempt to rescue a notion of analyticity in terms of that of necessity, but then considerations brought by Kripke have already convinced us that *all* true identity sentences are necessary, including those that show no logical or conceptual connection. So it would seem that if there is anything relevant to cognitive value in Frege's remark, it is the issue with the extension of knowledge. That is, the puzzle with respect to "a=a" and "a=b" is how we can explain that someone who already knows the first sentence would extend his or her knowledge by coming to know the second (again, in the context of Mill's semantic theory for names and the principle of compositionality).

Given the above, it now appears that the question of the differing

---

[66]As has been noted by Salmon (1986), Frege's puzzle is not specifically a puzzle about identity since it applies to non-identity constructions as well. For example, the two conditionals "If Tully was a Roman orator then Cicero was a Roman orator" and "If Tully was a Roman orator then Tully was a Roman orator", seem to pose similar problems.

cognitive values between "a=a" and "a=b" becomes the question of what an extension of knowledge in this context could mean. Well, here things begin to be a bit murky since Frege never told us in what an extension of one's knowledge would consist. One suggestion is that to extend one's knowledge is just to learn a new fact about the world.[67] If we suppose that this is the issue with Frege's puzzle, then solving it would amount to introducing an informational component into linguistic meaning, perhaps in the way of encoding pieces of information in linguistic expressions. Probably the *locus classicus* of this sort of interpretation of Frege's puzzle is in N. Salmon (1990:409):

> Language is primarily a medium of communication, particularly the communication of information...The Fundamental semantic role of a declarative sentence is to encode a piece of information...A declarative sentence may be said to contain the piece of information it encodes and a piece of information may be described as the information content or cognitive content of the sentence. The piece of information encoded by a sentence is what philosophers generally mean when they speak of the "proposition" expressed by the sentence.

If we accept Salmon's informational line, then the question concerning an extension in one's knowledge turns into the question of what to make of the information content of a sentence. Intuitively, and put in the material mode (MM), the information content of a sentence or a subsentential expression

---

[67]Tascheck (1992) objects to this suggestion since he thinks that it is too close to Russell's theory. According to him, by an extension of one's knowledge Frege meant the logical appraisal of beliefs and assertions, a task which relies on their logical (i.e., inferential) properties rather than on their reference. In that Tascheck is closer to the 'formal mode' formulation of a difference in cognitive value given below. But then, if Tascheck thinks that such inferential connections between concepts are *constitutive* of their cognitive content, then his interpretation of Frege makes the latter into a genuine inferential role theorist and thus, indeed, would mark important differences between his semantic theory and Russell's. I think that this interpretation is strongly motivated especially due to Frege's treatment of empty names as having a sense but no reference, something I shall discuss below. But it is possible to adopt a different interpretation of Frege according to which inferential role is an *indicator* of identity or distinction in cognitive content, rather than constitutive of it. And since the true inferential role of concepts can only be ascertained by an *ideally rational* agent, then one opens the door here for rational psychology to tell us what these relations are. I will return to this point below, but mainly in chapter Four section 8.

consists of some representational property in virtue of which it would be about certain objects and/or events in the world on their modes of combination or connection to each other. Put in the formal mode (FM), the information content of a sentence is that which a linguistically competent and ideally rational speaker of the language would come to believe to be the case *about the world* upon coming to understand the sentence. From this we can derive a notion of a *difference* in the information content of a sentence or an expression as follows:

Material Mode: Two sentences A and B would differ in their information content just in case they would represent different things in the world, or the same things differently connected.

Formal Mode: Two sentences A and B would differ in their information content just in case a linguistically competent and ideally rational speaker of the language who already knows about the things and modes of connections mentioned by one of them would come to know of connections or things in the world one was previously ignorant about.[68]

Let us assume, for the sake of the discussion, that Salmon's informational interpretation of Frege was right on point and continue as if the issue is to explain informational differences between sentences which are putatively about the same states of affairs connected in the same way. That is, the question now is how can "a=a" differ in its information content from "a=b" given that both are true and hence that both are putatively about the same objects connected in the same way (by identity in this case). Frege's way of doing so went via the introduction of senses.

3. Frege's Solution to Frege's Puzzle

Frege's solution to the puzzle was by introducing a distinction between two meaning relations: Sense (*Sinn*) and Reference (*Bedeutung*). Frege claimed that every primitive symbol, and any well formed formula which does not contain free variables, is a name which *expresses* a sense, and ideally denotes an object,

[68]In fact, some qualification about epistemic ordering is required here since coming to know of "a=a" after knowing "a=b" would not extend one's knowledge in the way defined above. So I think the point should be made that sentences differ in their information content if on at least one of their epistemic orderings the differences implied in the text would come about.

the reference.[69] So far we seem to be in the ball-park of the two-factor theory. The differences come about in Frege's claim that the relation between the two semantic relations is that the sense of a name is the mode of presentation of its reference (Frege 1892b). Thus, e.g., proper names such as "Hesperus" and "Phosphorus" might expresses distinct senses which still determine the same reference - the planet Venus.

The solution to the puzzle now depends on senses also having a cognitive value, or what was suggested in the last section, information content. For this purpose, Frege seems to have implicitly adopted the following criterion, which I will dub *Frege's Criterion:*

For any two (Fregean) names N1 and N2, N1 and N2 differ in their cognitive value (or information content) iff N1 and N2 express distinct senses.

This principle then allows for the *explanation* of a difference in cognitive value of two terms to be given by reference to a difference in their expressed senses. Alternatively, it implies that an identity in sense is sufficient for an identity in cognitive values. But whichever way you want to put it, it now seems that the difference between sentences which involve coextensive terms such as (1) and (2), or

(6) Aristotle was wise,
and
(7) The Stagirite was wise,
is in that they express different information contents. This means, as we saw, that they either represent different things in the world, or that they represent the same things differently connected.

Let's take a closer look at (6) and (7). We know that the object about which both names are is the same, and there is no difference in the connection between whatever they name and the property of *being wise*. So where can the informational differences come from? Well, in a footnote to his discussion of senses, Frege says that the sense of "Aristotle" might be "the pupil of Plato and Teacher of Alexander the Great". Namely, Frege seems to give the sense of "Aristotle" in term of a descriptive content, a cluster of properties, something

[69]Formulas which contain free variables are either predicates or relations. This includes such abstract nouns as 'a person', which for Frege should be construed as the open sentence 'x is a person'. Unlike names, the meaning of predicates is not an object but the concept (an incomplete object) that they express. Apparently Frege thought that predicates also have a sense, but the issue of what the sense of predicates might be is even more obscure than the one concerning the sense of names.

whose semantic relation to its denotation is that of *satisfaction*.[70] Then, presumably, the information content of the name could be given by reference to this cluster of properties. The question is, though, whether Frege thought that this is what the sense actually *consists* in, or that this is just a way of *expressing* it.

Under the influence of Russell's theory, and more lately due to Dummett's interpretation, it is common to think that the former is what Frege had in mind.[71] And perhaps some motivation for that interpretation can be gathered from the issue of empty names. For here it would seem that the only way to give the sense of a name is by reference to some such description. This is because definite descriptions seem suitable for the role of expressing a cognitive content even if they are not satisfied by an object, in terms of the properties they express. Thus, when Frege says that the thought expressed by the sentence "Odysseus set ashore at Ithaca while sound asleep" remains the same whether "Odysseus" has reference or not (1892b:63), it seems that it is some such unsatisfied descriptive content which he has in mind that supplies the cognitive content of the name.

But in fact, this consideration regarding empty names does not conclusively tell in favor of the descriptive theory. I will now try to show that there is a referential alternative to the descriptive theory which is consistent with a solution to Frege's puzzle and the issue of empty names. Thus, it can be claimed that there are other ways to make senses independent of reference other than through the descriptive theory; for example, via a modal conception of truth-conditions. Under such a conception, although "Odysseus" does not have an actual reference, it can be said that it has a *possible* reference, one which makes the sentence true (or false) in a counterfactual situation. In this case, as well, the thought would remain the same even if the name lacks an actual reference, for the idea now is that the *meaning* of "Odysseus" might just be that possible reference. That is, the claim could be that the reference of the term does not have to exist for it to have a meaning. As to the cognitive value of the term, indeed, if we adopt the informational conception, then unless the object is instantiated the name would lack such a value since information can only be about that which exists.

In passing, it should be noted that these last considerations are consistent

---

[70]A description 'the x(Fx)' is satisfied by an object *a* iff *a* and only *a* has all the properties expressed by Fx.

[71]Dummett (1973) thinks that this descriptive content need not be expressible, but he still takes it to be something like a criterion for identifying the object named. For a non-descriptivist interpretation of Frege, see McDowell (1977, 1986:143-4).

with the apparent failure of identity sentences involving non-referring terms to express an information content which is not otherwise intentionally implicated. For example, the putative identity sentence 'Hamlet is the prince of Denmark' does not convey any information about the world, except that concerning the prosaic structure of the Shakespearian tragedy. But then, this is a piece of information which makes essential reference to other minds (that of Shakespear in this case). On the other hand, that identity sentence surely has a meaning, as a function of the meaning of its constituents, although not a cognitive value. Again, this is consistent with the theory where the *meaning* (i.e., reference) of empty names (including false sentences) is an object or state of affairs that are only possibly instantiated, and where the cognitive content is the object or state of affairs in actuality.

Frege, of course, says nothing of this kind, but if he had anything like the informational view of cognitive content in mind, this is what he *should* have said. That is since otherwise, there is no possibility of taking the cognitive value of empty names to consist in the their descriptive content, if that descriptive content does not convey information. But the descriptive content of empty names is that which is not actually satisfied by definition. This implies that empty names have not only no reference, but no cognitive content as well.

The considerations I presented in the last few paragraphs in favor of a non-descriptivist solution to the problem of the meaning of empty names can be now generalized to names in general. My suggestion is, then, to assign to each term in a language both a meaning (i.e., reference) and a cognitive value, the difference being that a term has its meaning essentially (that is, in every possible world), whereas its information content is had by it only in those worlds in which it exists with the object, or in which it exists with its corresponding state of affairs.[72] But this now shows that we no longer need a notion of the descriptive content of a term as that which is satisfied by an object, or that which fails to be satisfied (in the case of empty names). Instead, we can switch to talk about actual or possible instantiations of properties.[73] As we shall see, this is more or less the

---

[72]This implies that terms for impossible objects or states of affairs lack both kinds of meaning. For example, this theory implies that "the greatest prime number" has no informational value, at least not as a singular term. Surely each of its constituent terms has one, but the point is that we would fail to make a significant assertion by using that singular term. I think this latter idea is still consistent with Frege's context principle (in Frege 1884) as interpreted by Dummett (1973), but I lack the space here to extend on it.

[73]Indeed, the difference between the two ways of talking about cognitive contents is only that of getting rid of the criterial conception of senses, as ways

line which informational semantics has taken[74]. But while information theorists have concentrated on what I will call 'general properties' (probably with the exception of Fodor 1987), I suggest that we also entertain the possibility of informational relations to 'singular properties', such as *being Hesperus* and *being Phosphorus* (or alternatively, that of *Hesperizing* and *Phosphorizing*). In which case, the differences in cognitive values between the two names would consist of differences between instantiations of distinct singular properties. As is well known, the idea that the two terms of identity sentences might have been distinct terms is not very popular these days (Cf. Kripke, 1972; Putnam, 1975), but I will not try to defend it here.[75]

Now some would say that all that talk about the property of *being Phosphorus* as distinct from that the property of *being Hesperus* and the like is just more evidence that Frege's conception of cognitive value could not have been informational in the first place, and if anything, that this is more true to Russell's theory. So I suggest that we give a look to Russellian semantics and then see (in section 5) what we want to preserve from both theories.[76]

4. Russell's Solution to Frege's Puzzle[77]

The major difference between Russell's solution to Frege's puzzle and that of Frege is that Russell located the differences in cognitive values of coextensive

---

to determine reference. Otherwise, one can take talk of descriptions and talk of properties to be on a par.

[74]But not where Salmon had taken it. For Salmon, the differences in information content are pragmatic, rather than semantic (Salmon 1986). But that is surely *not* Frege's theory.

[75]Kripke, and possibly Putnam, would also object to the idea that a term can have a meaning which is not its actual reference. According to my theory, this is to confuse meaning with cognitive value.

[76]An alternative suggestion, one which does make sense depend on *actual* reference, is McDowell's theory (1977). McDowell then claims that thoughts involving empty names are not real thoughts but illusions. McDowell then proceeds to claim that what would explain our behavior in the case of empty thoughts are mistaken second-order thoughts to the effect that we entertain the first-order thought. But I wonder, if we do not entertain the first order thoughts, *if they do not exist*, how can we have second-order thoughts *about* them?

[77]The discussion here relates to Russell's post 1905 theory.

terms in a unified notion of meaning which he identified with their referential truth-conditions. In contrast to Frege's Sense-Reference theory, where each semantically relevant expression had two kinds of meaning relations, Russell has advocated only one meaning relation, a referential meaning relation, which was supposed to serve different functions for different types of expressions: In the case of syntactically simple linguistic subjects, their meaning was the object they named. In the case of syntactically complex expressions and linguistic predicates, their meaning was the property or proposition that they expressed. This unified meaning relation is then used to connect the cognitive values of expressions to their truth-conditions. in the following simple formula, which I will call *Russell's Principle*:

Two sentences A and B differ in their cognitive value only if A and B differ in their truth-conditions.

What were the constituents of truth-conditions for Russell? To answer that we need to start with the Russellian semantics for names.

Consonant with the general informational view of cognitive contents, the cognitive content (meaning) of a name should be the object it refers to. From which it follows that if two terms differ in their cognitive content, then either they do not refer to the same object, or they are not really names. Rather, they show themselves to be 'denoting expressions' - definite or indefinite descriptions whose meanings are the properties they express. Thus, for example, if a name such as "Phosphorus" produces instances of Frege's puzzle together with "Hesperus", then this shows that these are not genuine names and should be analyzed via their associated descriptions, encoded as 'The F' and 'The G' respectively. In this way, the cognitive values of ordinary names, and hence their referential semantic values, are given in Russellian semantics in term of the semantic values of the *properties* which the descriptions express, where properties are construed by reference to propositional functions.[78]

---

[78]The idea that properties are propositional functions is that for each property there is an ordered set consisting of objects and a proposition. For example, the property *being wise* is construed as the propositional function (λx)(the proposition that x is wise) which takes an object, say Socrates, to give the proposition that Socrates is wise. However, this analysis would not go though if "Socrates" would prove to be not a genuine name but a covert definite description, say 'the θ', a description which might in turn consist in other proper names and propositional functions (as in 'the teacher of Plato', 'A citizen of Athens' and so on). The idea then is that in the *final analysis*, the proposition which is expressed by the sentence "Socrates is wise" will include only

Given this, we could now construe the sentence
(8) Phosphorus is wet,
as a conjunction of an identifying description and a propositional-function which represents the property of *being wet* thus:
(10) $(\exists x)(Fx) \& (y)(Fy \supset x=y)) \& Wx)$.
Similarly,
(9) Hesperus is wet,
would be represented as
(11) $(\exists x)(Gx) \& (y)(Gy \supset x=y)) \& Wx)$.

By giving this analysis, Russell in effect claims that the logical form of sentences which combine a definite description with a predicate are not on a par with sentences with genuine referring expressions, that is, those which do not introduce Frege's puzzle; for example, a sentence such as 'that is a wet planet'. This latter kind of sentences, according to Russell (1956:247) expresses a proposition in which the object being referred to figures as one of its constituents. Such propositions, which were later dubbed by Kaplan (1978) as 'singular propositions', owe their cognitive value to the existence of the object of reference, and thus can be said to be "contextualized" in that their assignment of truth-conditions will depend on whatever object it is that happens to be picked up by the context of utterance or thought. Still, it is just that object on which the cognitive value of the sentence depends.

As opposed to this, sentences involving ordinary referring expressions, and by extension, definite descriptions, have no such contextual dependency, but at the price of a loss of reference. The important point to notice about this analysis is that while the analysand expresses the fact]that some object uniquely instantiates the cluster of properties expressed in the description, any *mention* of the instantiating object has dropped out. That is because the object, which putatively satisfies the description, has been *quantified over,* which means that it can be just any object. This is not the former case where the context is guaranteed to supply an object as a cognitive value, but one where whether the context supplies such an object or not makes no cognitive contribution to the expression.[79] The result is that whatever differential informational contribution

---

propositional functions and variables. Of course, all this needs to be expressed in the formal mode, namely by reference to logically proper names and descriptors. See below.

[79]We would be gaining some information if we knew that some object, rather than none, satisfied uniquely the description, but that is to know the *truth* value of the sentence, not its cognitive value.

ordinary names make in Russellian semantics, it will be *in the properties* they express, not in the objects to which they putatively refer. As far as such names are concerned, then, their Russellian truth-conditions are purely universal.

I think that we can already see at this point that Russell's truth-conditional theory comes quite close to my informational emendation of Frege's theory, aside from the fact that Russell has pretty much eliminated our common objects (tables, chairs, electrons) from playing any informational role in his semantic theory. But to be sure, this is something we should find objectionable. If the sentence that Phosphorous shines conveys any information for me, it is about Phosphorus, not about *some object or other* which satisfies the description we presumably associate with the name. Perhaps better, we should say that the cognitive value of 'Phosphorous' is that some unique object Phosphorizes. Here, again, there is no need for an identifying description, the work is done just by reference to the actualization of the relevant property, as opposed to the satisfaction of the identifying description.[80]

Now, it was already implied above that for Russell's solution to Frege's puzzle to succeed, it is crucial that each of the genuine names used in his semantic analyses be purely referential, that is, that the meaning of these names would consist just of their reference. This is required in order to avoid reintroducing Frege's puzzle at the level of the analysand, something which might happen if two names or descriptors would denote the same object while still affecting a difference in cognitive content. Such a case could arise, for example, if an analysis of 'Plato' was to be given in terms of 'the teacher of Aristotle', which can then recreate Frege's puzzle with respect to 'the teacher of the Stagirite', and so on. That is, it follows from this theory that we need to make sure that our descriptors and names are logically proper, which means that they bear a 1:1 relation to the (named) things and properties in the world. We can put this meta-linguistic condition as a constraint on Russell's theory, one which I will dub the Principle of the Logical Propriety of expressions (PLP):

Let RL stand for the language in which Russellian analyses are to be couched. Then,
For any expressions 'x' and 'y' in RL. If 'x' refers to, or expresses, X, and 'y' refers to, or expresses, Y, then if X=Y, then the cognitive value of 'x' is the same as that of 'y', and 'x' and 'y' are logically proper terms.

Put in terms of the formal mode definition of cognitive values, the principle of logical propriety implies that any linguistically competent and rational speaker

[80] Cf. my remark above about the redundancy of the descriptive component also in the case of Frege's senses.

of the language who would come to believe what 'x' expresses would also come to believe what 'y' expresses, and vice versa.

How did Russell seek to guarantee the satisfaction of that principle? By introducing another principle, that of acquaintance.

According to Russell, the cognitive values of our sentences and thoughts are things with which we are directly acquainted, where being directly acquainted with an entity meant that one can immediately recognize which object it is, and hence, whether it is identical or distinct from another entity. Thus, if one is acquainted with X and Y, and X=Y, then one could not fail to know that fact (cf. also Tascheck, ibid.).

As we know, however, taking this route had dire consequences for Russell's theory. For one, and this is akin to a point we have seen earlier, it constrained the kind of entities that could serve as the cognitive values of our thoughts and sentences to only those to which presumably we have some privileged epistemic access - sense data, the self, and a special class of universals. This then had the implication that it is not, after all, the worldly objects and events that our thoughts and sentences are about, at least not the objects and events which take so much of our interest and on which we depend for our existence. To echo the words of McDowell, we are again at a risk of "losing our world".

The other problem with this 'epistemic return' of Russell's theory is that it blocks the possibility that what one means by one's terms is independent of that with which one is acquainted regarding its presumed reference. For example, it was argued by Kripke (1972) and Donnellan (1970) that people could be said to refer by a name to a concrete or an abstract object (such as a kind) even if they are 1) ignorant about all of its uniquely identifying properties, or 2) mistaken about them. That is, the claim was that being acquainted with a set of properties corresponding to the identifying description of, say, 'Moses' is neither necessary nor sufficient for reference to the Biblical figure (assuming there was one). This is because we would be referring to him, so it seems, even in the case where we are in possession of a wrong or incomplete set of such properties, and still not to any doppleganger of his. This argument against the descriptive theory was then used to motivate the claim that names refer to their reference not via a descriptive component but "directly".[81] The idea was to return to a non-descriptive conception of reference in which a suitable contextual relation between the sentence (or thought) and reality would make for its cognitive value.

However, this shift to the "extra-intentional" that was taken by proponents of the theory of direct reference has been objected to on the grounds

[81]Other forms of argument against the descriptive theory are the modal and semantic arguments which I will not discuss here.

that it implies a "transcendental standpoint outside our present set of representations and their object (Rorty 1980:293). Indeed, it seemed that underlying the anti-descriptivist stance was the assumption that having false beliefs about an object stands in the way of reference to it, something which would threaten our very touch with the world. The remedy offered, then, was of making for the semantic connection to the object from "without" (say, by reference to causal and sociohistorical relations), independently of the supposedly fragile epistemic capacities of the mind that have given rise to these skeptical conclusions. But we should note, with McDowell, that these very same epistemic presuppositions underlying the theory of direct reference are no less involved with the Russellian principle of acquaintance and the assumption that "the intentional nature of a thought could determine...that it was about a certain object only in the manner codified in the descriptive model." (McDowell, 1986:162) That is, it would seem that both Russell and the anti-decriptivists have bought into the conception that intentionality could fix a reference only via the theory of descriptions, as in the theory from 'On Denoting', or alternatively, via the acquaintance principle. According to McDowell, though, once we go beyond this conception, we allow for the possibility of intentional thoughts which will be directed to the objects that Fregean senses determine independently of any Cartesian underpinning, just because senses are already essentially object-dependent.[82] In which case, there would not be a need for *fixing* a reference through descriptions, since the objects are already given in their senses. And there would not be a need to secure PLP via Russell's acquaintance principle, since senses contribute their own enriched grain to the proposition, hence the solution to Frege's puzzle.

In my view, there is an insight in McDowell's exposition which I think we should adopt, this is the idea of the dependence of thoughts on their reference. However, I disagree with McDowell in that he takes senses to be dependent on the *objects* they present (1986:142), whereas, in my view, their dependence should be on the instantiation of a finer-grained set of properties *in the object of reference.* Because McDowell takes senses to depend on their objects of reference, their cognitive contributions to any thought would have to be as coarse-grained as that dependency allows. But to be sure, this conception of object-dependence could not account for the *differences* in cognitive values between coextensive terms . For the only way to make for such differences is to make senses depend on differences in *truth-conditions.* But again, this is Russell's

[82]That senses are object-dependent is shown, according to McDowell, in that the thought would lack a truth-value in the absence of the object. In which case, that is, where there is no object, it is not the same thought which would be expressed but another one (see also McDowell 1982).

theory and not Frege's, even on McDowell's interpretation of Frege's theory.[83]

So my claim is that McDowell came quite close to presenting the sort of theory I think could solve Frege's puzzle of cognitive values but just shied away from identifying the differences in cognitive values of *ordinary* terms with differences in their truth-conditions. In fact, in my view, what is missing in his solution is just this further step of identifying Fregean senses with finer-grained truth-conditions, with the additional imposition of some form of Russell's principle to block the puzzle from reappearing. In my view, again, all we need to do in order to secure something like the principle of logical propriety for ordinary terms, that is, to make sure that differences in cognitive value of such terms imply differences in their truth-conditions, is to make thoughts *supervene* on truth-conditions, and then add certain qualifications so that the same thought could not be about distinct conditions. As we shall see, such a model of the dependency of thought on truth-conditions is built into Fodor's causal theory of content which I shall discuss in chapter Six. But before I end this section, I want to make the following remarks vis-a-vis the descriptive theory, since in my opinion, descriptivism has a grain of truth in it which I think we should preserve.

Suppose that we take the descriptive component of a name to consist in a mode of fixing its reference, not as constituting its semantic relation to the object, but as *subordinate* to it (in the sense that the semantic relation can be determined independently). That is, the idea is that we take the mode of reference fixing as an inessential means to secure the semantic value of a token expression on a particular occasion of its use (including tokens of names and perhaps even indexicals), and thus replaceable in principle by any other means successful as it is. For example, we might presume that in a world in which Aristotle was the teacher of Alexander the Great, using that description would enable the users of tokens of the name, we should presume, to affect a reference fixing, just as Frege and Russell thought (and indeed Kripke too). But now the difference is that we allow that for the same purpose *any* other means could be used, *including false descriptions and theories which incorporate misleading or missing information,* provided they issue in the same result: Fixing the right reference. For to be sure, the issue is not whether the description or theory is true,

[83]Well, Frege came quite close to this in the Grundlagen (1882/1964:90)) where he expresses himself to the effect that the thought that a sentence expresses depends on its truth-conditions. If the truth-conditions of the sentences are different, then the thoughts they express are also different. In other words, here Frege takes identity in truth-conditions to be necessary for identity in the corresponding thoughts. Of course, this still leaves it open that the thoughts would be different while the truth-conditions are the same, as a solution to the puzzle requires.

and hence whether it was semantically necessary or sufficient to fix the reference, but whether it can *reliably* do so, however this comes about. For another example, arguably, the ancients were able to refer to Venus although most of the information they had about the planets was false, and arguably again, it was their *false* theories which enabled them to accomplish this, at least to some extent.[84] Still, this does not establish the description theory of *content*. That is because we would not want to think that some or many of our *concepts* are stocks of false or incomplete information since a) that would reintroduce the problem from meaning incomparability discussed in the last chapter as it would be hard to come-by two such concepts with the same content[85] and b) their theoretical content would become too promiscuous to count as semantic content. But then, this is surely not 'direct reference' either, since we now allow that reference to an object can be established via some description or other, including false or incomplete ones.[86] Indeed, this shows why the theory of direct reference should be distinguished from the referential truth-conditional theory. For we see that one can advocate a theory in which the reference of a name is secured via a description, although its cognitive value is still its reference.

To conclude then, it now seems that Russell was right to think that the constitution of content requires quantification, but his mistake was, I argue, that he quantified over truth-conditions instead of the modes of their fixation. Similarly, we can see that direct reference theorists were mistaken in thinking that false beliefs and theories cannot secure a reference, a mistake which resulted in what I called the extra-intentional theory of cognitive value. But my claim is that cognitive value is an issue internal to intentionality, even if such cognitive

---

[84] This in the same way that one can build an accurate thermometer while using a false or an incomplete theory.

[85] We should just imagine how many ways there are to get a theory wrong.

[86] It might be that this is what Donnellan had in mind in his notion of the referential use of descriptions, as in 'the murderer of Smith' which can still refer to the accused whether or not he was the actual murderer of Smith (in Donnellan 1966). Surely in this case (as well as in the other examples Donnellan gives), the reference to the accused was fixed via that very description, even if it was not via its identifying descriptive content. Rather, according to Donnellan it is the "context" which makes for reference. This I find very vague and also problematic with respect to the issue of grain that all contextual/historical theories of reference face. According to my theory, it is the existence of a certain regularity between uses of the name and occurrences of the object of reference that determines the content of the term.

value consists in "just" extensions. In my view, the way to make the extensional a part of intentionality is not by introducing intensional entities that would present it to the mind, like Fregean senses, but by making extensions, that is, truth-conditions, richer in content with respect to human systems than they are with respect to non-intentional systems. In other words, we can say that intentionality consists in a fine-grained access to the world, not in a mediated access; and a theory of intentionality should be a theory which investigates the nature of that fine-grained access. These issues will be further explored in the following chapters.

Chapter Four

Propositionalism, the Dyadic Theory of the Attitudes, and a bit more on Frege's Puzzle

In the last chapter I discussed the issue of the cognitive value for the semantics of natural languages where I presupposed that this was correlated with the issue of the cognitive values of mental states. It is time to be more explicit and specify how that discussion bears on the nature of intentional mental states. Now, in chapter One I briefly mentioned the relational theory of the attitudes under the representational theory of mind. We saw there that according to RTM, propositional attitudes are construed as a ternary relation between an organism, a mental representation, and a proposition. In fact, this conception of the mental is an extension of the theory according to which intentional states are construed as a *dyadic* relations between organisms and propositions. Underlying that conception are certain formal, semantic, and psychological considerations which were subjected to various forms of criticism ever since the inception of the theory some hundred years ago.

Perhaps the most controversial aspect of the dyadic theory is its claim that intentional states, as psychological active states of the organism, involve what seems to be a mysterious relation to Platonic entities, that is to senses or propositions. What is more, given our commitment to intentional realism, the idea is that it is this relation to propositions which gives these psychological states their causal powers. In any event, I will have to say very little in this chapter about the nature of the relation to propositions, although what I have to say will show that once we get rid of the occult epistemological terminology usually associated with that relation, like the notions of 'grasping' or 'entertaining' a proposition, life with propositions can be much easier.[87] My issue with propositions here will be different. In the main, I will rehearse some of the better known arguments why relations to propositions cannot account for the sort of

[87]I already touched upon the nature of that relation when I discussed CRT in chapter Two. There the idea was that the nature of that relation is functional. By going functionalist, a theory can also give an answer to the question of the causal powers of mental state, since the relation to a proposition is defined by reference to causal powers. Due to the failure of CRT, though, a different account of the relation between the causal roles of mental states and their semantic content will have to be given. This I shall attempt to do in the next chapter.

intentional relation that we need for psychological theorizing, and then also add a few arguments of my own. In particular, the point will be that propositions cannot supply the level of grain that intentionality requires, since they are logical entities, closed under logical transformations. Thus, from the point of view to be presented here, to answer the question of what makes a mental state intentional is to show how to fine-grain the content of such a state, in the way shown in its empirical manifestations.

The road I intend to take to arrive at that conclusion will be somewhat long and winding. After having established in the first section that the attitudes should be construed as relations between agents and worldly objects and states of affairs, we shall take a look at a particular theory about what makes the intentional relation be about the objects and states of affairs it presumably represents. The theory, which I call propositionalism, is one where the attitudes are construed as relations between agents and a token psychological state (of theirs) which *refers* to a proposition. To show how I arrive at this theory, I will present in section 2 a well known formal argument to the effect that ascription sentences should be analyzed as expressing *the attitude relation* between an agent and the semantic value of the 'that'-clause of that sentence. Contrary to what some might expect, I will not argue for this linguistic move but merely use it to derive, in the section to follow, the claim that the semantic objects of 'that'-clauses are propositions. The reason why I think that we can do without a justification for this reification will become clear when I show that there is an alternative to propositionalism in the form of a theory where agents can be said to instantiate representational *properties*.[88] However, my strategy will be that so long as the relational theory can be held coherent, given the formal considerations, it is better to hold to it rather than not.

Now Propositionalism, which is probably the most (in)famous position in the philosophy of mind in the last 100 years or so, mind, has been traditionally associated with the semantic theories of Frege and Russell. But in section 3 I will belabor some of the arguments which show that neither Frege's semantic theory, nor Russell's, strictly entail propositionalism - at least not if by propositions we understand entities which are more fine-grained than truth-values. That Frege's theory does not entail propositionalism is shown by an argument called 'the slingshot', according to which all sentences with the same truth-value refer to the same thing. That Russell's theory does not entail propositionalism is shown by

[88]That is, the idea is that if both the relational theory and the instantial theory are as well motivated by the same argument form, and given that in any case, this argument form is not intended to support intentional realism as such (since this one is motivated independently), then we can treat this argument form as a heuristic device.

the fact that 'that'-clauses, according to that theory, are not referring expressions. However, my claim will be that the referential aspect of 'that'-clauses can be rescued, at least from the claws of the slingshot, by countering a crucial assumption in the argument. This is the assumption that logically equivalent sentences express the same proposition.

Now, in chapter Three I suggested that we should construct a hybrid theory of propositions which will be as fine grained as Fregean senses are, but still representational as Russell's truth-conditional semantics is. In section 4 of this chapter I will have an opportunity to review a theory which promises to come up with a similar picture, by individuating propositions by reference to a fine-grained ontology of properties and relations. This is Bealer's theory of PRPs (or Properties, Relations, Propositions). Basic to this theory is a distinction between two conceptions of entities, one concerning worldly entities and the other concerning those constructed from worldly entities via special logical operations to give concepts. The claim of the theory is that only in entities of the second kind do we arrive at the degree of grain that propositions require. As we shall see, this theory faces some technical difficulties and in addition, it seems to renege on its promise to give a fine-grained *truth-conditional* semantics for propositions. Then, in section 6, I will turn to some meta-linguistic approaches to propositions. Here, as well, my claim will be that such a solution cannot supply a viable cognitive conception of a proposition.

At this point, my conclusion will be that given this somewhat comprehensive survey of propositional theories, the failure to find a psychologically adequate conception of propositions shows that there is something inherently wrong with propositionalism. In section 7 I will claim that propositions cannot be psychologically adequate since they are closed under the logical, which is a constraint too restrictive for psychological states. Then, in section 8 I will offer a non-logical, empirical individuation scheme for the attitudes, based on Mates' famous test for sameness of meaning. My claim will be that we can extract from Mates' test an individuation scheme for the attitudes which is as fine-grained as the empirical data suggests since it is based in that very empirical data.

1. Intentionality and Representation

According to Brentano's famous characterization of the mental,

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon

> includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgment something is affirmed or denied, in love loved, in hate hated, in desire desired and so on.
>
> This intentional inexistence is exclusively characteristic of mental phenomena...Consequently we can define mental phenomena by saying that they are such phenomena as include an object intentionally within themselves. (Brentano 1874, in Chisholm 1960:50)

It was claimed by Chisholm (1957), that Brentano's theory of intentionality is a thesis about a *relation* of a special sort between mental states and objects in the world, not about a special kind of "inexistent" *object:* objects like sense data, ideas in the mind, or universals in direct acquaintance. This relation is special in that it would still obtain even when the worldly object did not exist: Diogenes could have been looking for an honest man even though there might not be any such man. Similarly, Diogenes could have believed that there are some honest men, although that does not entail either the truth or falsity of the statement that there are some honest men.

Whether or not this interpretation is true to Brentano's passage (I actually find this difficult to believe) is not the issue in which I am interested here. Rather, the point is that if, indeed, intentionality is a relation between the mind and worldly affairs, then it is possible to construe it as a *representational* relation. If so, then the question of intentionality becomes the question in virtue of what mental states are representational?

In general, several options have been suggested here. One, which is also the classical answer, is that mental states are representational in virtue of their propositional content. For example, it was claimed by John Searle that

> A belief is a propositional content in a certain psychological mode, its mode determines a mind-to-world direction of fit, and its propositional content determines a set of conditions of satisfaction (Searle, 1983:15).

Searle's definition of belief, which we can also take to cover more generally all intentional states, raises two issues with respect to such states that bear emphasize in this context. One issue is the psychological one which bears on the question of *what makes* mental states intentional, which concerns the structure of the mental. I will say a few things about this below, but it will not be the main topic of discussion here. The other issue is the semantic one, which will occupy us in the rest of the chapter. What I have to say about it now will serve as a sort of compass for the discussion that lies ahead.

In chapter Three, when I discussed the question of the representational properties of mental states, I tied it down to the issue of their cognitive values. We saw there that for mental states, or linguistic terms, to differ in their cognitive

value is for them to differ in their information content. Then it was claimed that this difference in information content can obtain even in the case of states which are *about* the same object or states of affairs. This point can be best appreciated in the context of truth-conditional semantics.

According to truth-conditional semantics, a thought is about that object and that property which together would make the thought true. E.g., the thought that *a* is F would be about *a* if the thought is true iff *a* is F. We already saw (in chapter Two) that a theory of meaning based on such semantic equivalences was given by Davidson as based on Tarski's convention T sentences. However, given my adaptation of Russellian semantics, we should mention that we advocate a finer-grained intentional relation than the one found in Davidson's theory. For example, according to Davidson's theory of meaning, the thoughts corresponding to the sentences 'Cicero is emulated' and 'Tully is emulated', are given the same truth-conditions. But this clearly fails to supply a notion of content that satisfies the empirical constraints where agents can be found systematically to have the first thought but not the second. According to my proposed theory based on a theory of informational content, the two thoughts should be given distinct truth-conditions and hence distinct intentional contents. This would make 'Tully' thoughts have a different reference (or express a different *property*) than 'Cicero' thoughts, which in effect means that each of these terms would have a distinct content.

In line with my discussion above, I take it as a general semantic constraint on any psychologically adequate truth-conditional semantics that no two cognitively distinct proper names can have the same meaning (i.e., reference), although I claim that they can be *about* the same object. Similarly, no two cognitively distinct predicates can express the same properties, although they can be *about* the same collection of objects.[89] For example, according to this semantic constraint, although I allow "Tully" and "Cicero" to be about a certain individual, 'Tully/Cicero', I take the cognitive values of the two names to consist in the instantiation of two distinct properties, that of *being Tully* and that of *being Cicero*. More generally, I distinguish the object *about which* a term or semantically (e)valuable state is, from the referential content of that state; and I distinguish the state of affairs *about which* a sentence is from that sentence *'s*

[89]This implies that translations of names and predicates, as well as definitions, can have the same content only where they do not introduce cognitive differences. As we shall see below, since the question of when two terms show cognitive differences has to be settled empirically, questions of successful translations and definitions will have to be decided accordingly.

truth-conditional content.[90]

Let's call the semantic motivation for the assignment of content based on the issues raised in the last few paragraphs, the motivation *from without*. That is, the idea is that given my acceptance of the informational theory of cognitive differences, and hence of content, the external assignment of semantic properties to mental states and linguistic items follows in the way specified above. But surely there is more than this to the semantic valuability of mental states. For unless we assume that the relation between mental states and their semantic values is purely formal, then we would also like to know what is it about these states *that makes* them have those semantic values. In which case, the question of intentionality becomes the question *from within* as to what it is about mental states which makes them be about certain objects and certain states of affairs, rather than about others.[91]

I should note that this question, about what makes mental states intentional, is a psychological question because it requires us to say something about the attitudes themselves rather than about their semantic valuability. It is still a psychological question even if it is answered by reference to a relation between an agent and a proposition as an independent and abstract entity. For even on this theory, which I dubbed propositionalism, the answer is that mental states are intentional in that they are capable of establishing a relation of a special sort with things abstract. Indeed, the distinctiveness of this sort of relation is shown in the presumption that when we 'grasp' a proposition we can do so in only one way. I should note that the whole point of grasping a sense/proposition would be lost if we could relate to these entities in the same way that we relate to concrete objects. Concrete objects can have different aspects to them, hence different ways of thinking about them, hence the need for senses (on Frege's theory).[92]

In the rest of this chapter I will pretty much ignore the psychological issue about intentional states, since my interest here is to see if there is any acceptable

---

[90]This will require that we take a second look at what *aboutness* is. To anticipate, the idea is that if 'a' refers to *a*, and 'a iff b' is nomologically necessary, then 'a' is *about b* as much as it is *about a*.

[91]Cf. Wittgenstein's "bewilderment" regarding the phenomenon of intentionality, which he also put in the form of the question: "what *makes* my image of him into an image of *him*" (1953, p. 177. First italics mine).

[92]Indeed, by ignoring the issue of the nature of the *psychological relation* to senses and propositions, Frege has saved himself the trouble of dealing with the regress of senses that seems to be generated here. See also Fodor's discussion on the same topic in Fodor 1998, pp. 15ff.

propositionalist theory that can satisfy the empirical constraints on content specified above. This topic seems to be independent of the psychological one. Then, when I will claim in section 8 that propositionalism cannot, in any case, ground my theory of intentionality, the psychological issue of the relation to propositions will dissolve itself.

2. On the Logic and Semantics of 'That'-Clauses

The origin of the relational theory of the attitudes is in the logical analysis of attitude, or 'ascription' sentences, such as

(1) John believes that Hesperus is far.

The peculiarity of such sentences is in that they contain a sentence as a proper constituent, not by means of a boolean connective but through a 'that'-clause. As the following example shows, this is a phenomenon not unique to attitude sentences:

(2) It is necessary that all triangles have three sides.

The question is what to make of the logical form of such sentences containing 'that'-clauses. Now, there is a strong argument to the effect that 'that'-clauses should be generally construed as singular terms, with the intensional verb taken as either a predicate (as in 'x is necessary', 'x is believed', etc.) or as a relation. The argument consists in the claim that construing 'that'-clauses as singular terms is the only explanation for the logical validity of the following argument forms:

Leibniz's law is necessary
Leibniz's law is that identical items have the same properties
∴ It is necessary that identical items have identical properties

The claim is that if we construe the 'that'-clause as the singular term '{A}', we could formalize this argument quite neatly as follows:

□L
L={A} (an identity predication)
∴ □{A}[93]

[93] See Bealer 1982 and 1993b. Bealer notes that the higher-order sentential operator approach can be seen as a special case of the predicative approach. The only difference being is that the arguments would be taken in this case to range

A similar argument is used to show that 'that'-clauses occurring inside the scope of propositional attitudes verbs should also be treated as singular terms:

Whatever s believes is true	(x) (sBx ⊃ Tx)
s believes that A ;	sB{A}
∴ It is true that A	∴ T[A]

Another class of arguments which require representing 'that'- clauses as singular terms are those involving 'quantifying-in':

Shiran believes that Roy believes that A ;	aB{rB{A}}
∴ Shiran believes that Roy believes something	aB{(∃x)rBx}

Now, suppose we grant, for the sake of the argument, that the logical form of sentences involving 'that'-clauses has been decided in favor of the relational theory in whatever form (I will return to this shortly). Then the next question is about their semantic values. That is, given the principle of compositionality I discussed in chapter Three, it follows that 'that'-clauses, proper parts of ascription sentences, should make their own contribution to the cognitive content of those sentences. Here philosophers have diverged on whether these semantic values are particulars or universals. Nominalists think that the semantic values of 'that'-clauses can consist only of particulars or things made out of particulars: most notably, sentence-tokens in a public or a private language. Thus, Carnap's nominalistic theory is that 'that'-clauses refer to sentence tokens in a public language (in Carnap 1947), whereas Fodor's language of thought hypothesis is that the direct objects of 'that'-clauses are sentence-tokens in one's mental language (see chapter One). Of course, for such linguistic items to have a cognitive value, their own meaning needs to be specified. I have already discussed, and rejected, one such theory: conceptual role semantics. Below we shall have the opportunity to see whether the sentential or meta-linguistic approach fairs any better.

Opposite to nominalism there is the position according to which the objects of 'that'-clauses are universals rather than particulars. These can be either concepts or ideas in one's mind (mind-dependent universals), or mind-*in*dependent entities such as propositions. This latter position, which I called propositionalism, is perhaps the most prevalent, and at the same time controversial, in the philosophy of mind. We already saw Searle's commitment to propositionalism. Here is another example of a scholar whose allegiance to propositions is such that he is willing to base his entire theory on the attitudes on

the assumption that attitude ascriptions are what they appear to be:

---

over sentences (as in '□p').

> sentences in which a two-place predicate ('believes', 'says', etc.) connects two genuine terms, one typically simple syntactically ('Iago'), the other typically complex ('that Desdemona will betray Othello').This assumption...saddles us immediately with t-clauses as names of entities of some sort...that is...with propositions (Richard, 1990:5).

As is well known, there are some notorious consequences to this position, many of which concern the presumed abstract nature of propositions. Indeed, philosophers tend to complain that the problem with propositions is that they are Platonic entities whose nature is "very mysterious", and that they "can play no causal role in mind and language" (Devitt 1996:210). But in my view, the presumed abstract nature of propositions has received undue criticism in the literature. We can see this when we compare the disdain for propositions with the more "lenient" approach to properties that one encounters. After all, properties (or truth-conditions, or truth-values, or facts, or states of affairs) are also abstract entities and therefore one would say "mysterious", but still many of the foes of propositions accept them into their theorizing quite readily.[94] As to the causal efficacy of propositions, or rather the lack thereof, a point which has figured predominately in the arguments for the triadic theory of the attitudes (such as LOT and CRT, see e.g., Fodor 1978, 1980), then, again, I don't see much reason for that beyond the presumed abstractness of propositions.[95] But then again, properties are also abstract entities but still causally efficacious *when they are instantiated.* So maybe that is the issue, that properties are instantiated whereas propositions are not: They are not just abstract entities, but are more specifically abstract *objects.* And it might be claimed that although properties and kinds (of objects) get instantiated, objects as such do not. They either exist, or they don't.

In line with this sort of presumed "liberalism" with respect to abstract entities other than objects, Michael Devitt has claimed that attitudes should be construed not as a relation between an organism and a proposition, but between that organism and a token state which instantiates a certain representational

---

[94]Unless they are nominalists, in which case they replace them with no less mysterious entities: classes, possible worlds, not to mention that the concept of "a particular" is at bottom not much clearer than that of properties or propositions.

[95]And it cannot be because they are meanings either because this begs the question against intentional realism. Similarly, it cannot be because they are semantically evaluable entities since sentences, MRs, maps, mental tokens, are all semantically evaluable entities, but still, one presumes, causally efficacious.

property. This, he thinks, would give them the naturalistic seal of approval that propositions lack. For example, instead of taking Iago to be related to the proposition that Desdemona will betray Othello (in 'Iago hopes that Desdemona will betray Othello'), we should construe this attitude as a relation (functional, I presume) between Iago and a mental token which *instantiates* the property of representing Desdemona's betrayal of Othello.[96]

But still, there is the quantificational issue with attitude sentences as in

(3) There is something that both Desdemona and Othello hope,

or

(4) Othello doubts everything that Desdemona believes.

It seems that sentences like (3) and (4) imply reference to a common *object,* which might just as well be the propositional object of the attitude. Devitt's response here is inadequate, as I see it, since it invokes the notion of "synonymous tokens" (ibid., p. 213). That is, his idea is that believing or hoping the same thing is being related to synonymous tokens (cf. Davidson's idea of samesaying). But this would just make the issue of mental content depend on linguistic meaning, not something that intentional realists would endorse. Instead, I suggest that we construe (3) as

(5) There is a property of representing some state of affairs which both Desdemona and Othello instantiate in their "hope box".[97]

Similarly, we construe (4) as

(6) For every property of representing some states of affairs, if Desdemona instantiates this property in her "belief box", then Othello instantiates this property in his "doubt box".

I conclude this part of the discussion in that we now see that it is might be possible to do away with propositions as the objects of the attitudes after all. But the truth is that I'm still not sure how much this argument is conclusive against

---

[96]Or more accurately, of instantiating the property of *being a representation regarding Desdemona's betrayal of Othello.*

[97]This terminology of 'belief box' and 'doubt box' is a functional terminology, due to Schiffer 1981.

propositionalism, since it might be claimed that properties of representing states of affairs are just propositions. That is, we can now say that if at all, the above argument might be convincing against the propositional theory taking propositions as *objects*, but not against the theory taking propositions as abstract entities more generally (e.g., as properties). In which case, we might have to give-up on the relational theory, that is, of taking 'that'-clauses as singular terms, but not necessarily give-up on propositions. All it would take now is to bite the bullet and accept that properties can also be quantified over (that is, of propositions as representational properties), but who says that only objects can be quantified over? To think so, I would argue, is to show a nominalist prejudice.

Still, this is not to say that I don't think that there are problems with the idea of propositions as the semantic values of the attitude. As I claimed above, the problem lies, in my view, in that propositions cannot determine a notion of intentional content which is fine-grained enough to be empirically adequate. But before I get into that, I want in the next section to return to Frege and Russell's semantics, which for many is the locus classicus of propositionalism. As we shall see (although I don't pretend to make here any original claim), neither semantic theory strictly entails propositionalism.

3. Frege and Russell's Theories of the Propositional Attitudes

We saw that construing 'that'-clauses as singular terms seemed the best way to deal with certain argument forms. For Frege, this feature of 'that'-clauses just follows from that aspect of his theory which treats any semantically relevant expression as a name. But Frege's theory does not imply that the semantic values of 'that'-clauses should be propositions, but rather truth-values. The argument for this is called Frege's argument, or sometimes 'the Slingshot'.[98] In this section we shall see that the slingshot brings-forth another motivation for preferring the Russellian unified notion of meaning to that of Frege's semantic dualism.

The aim of the slingshot argument, as developed by A. Church (1956), was to show that any two sentences with the same truth-value must stand for the same thing.[99] The argument seems to presuppose the following assumptions:

(I) Logically equivalent sentences, that is, sentences true in the same models (or possible worlds), express the same proposition.

---

[98] As far as I know, the originality of the title is due to Perry and Barwise (1981) who wanted to capture the fact that the argument utilizes relatively basic means to generate such forceful consequences.

[99] For an extended discussion of the Slingshot, see also S. Neale 1995.

(II) The proposition expressed by a sentence determines its naming relation, i.e., its reference.
(III) The naming relation of a complex name is a function only of the naming relations of its semantically relevant constituents.
(IV) A proper definite description (The x)(Fx) refers to the unique object which satisfies the description Fx.

I will first give the argument and then return to discuss its assumptions.
The argument:
For each sentence *S*, let Ds be the definite description:

The number that is 1 if S and 0 if not-S. Then *S* will be logically equivalent to Ds=1, if S is true, and to Ds=0 if false.
Suppose now there are two true sentences א & ב (read: Aleph, Bet). Then:

A) א (assume)
B) $D_{א}$=1 (by logical equivalence)
C) Ref(א) = Ref($D_{א}$=1) (by I and II)
D) Ref($D_{א}$) = 1 (by IV)
E) 1=1
F) $D_{ב}$=1 (by IV)
G) Ref(ב) = Ref($D_{ב}$=1)(by (I) and (II))
H) Hence: Ref (א) = Ref(ב).

Thus, given assumptions I-IV, the argument shows that א and ב have the same Reference. Similarly in the case where both א & ב are false.

Now this argument, if sound, would not only have the consequence that all sentences with the same truth-value stand for the same thing, the True or False, as Frege put it, but that they cannot stand for (i.e., name) anything more fine-grained than these. That is, this argument seems to have the consequence that sentences cannot *name* propositions, states of affairs, situations, or most importantly for us, pieces of information. This is because all these entities are assumed to be finer-grained than truth- values.

Consider the report sentence "Copernicus believed that the apparent motion of the Sun is produced by the real motion of the earth". Suppose it is true. Then, if all true sentences have the same (referential) semantic value, we could replace the subordinate sentence in the attitude report by any other true sentence, for example, 'Bill Clinton was the 42nd president of the USA'. But surely the sentence resulting from this substitution is false: Copernicus had no beliefs about

Bill Clinton.[100]

Frege, of course, recognized this problem with his theory and hence made the stipulation that sentences in contexts of "indirect discourse" do not have their ordinary reference, but instead refer to their customary sense: the proposition expressed by the sentence. This was in fact how propositionalism was born. But clearly, that claim, that the semantic objects of 'that'-clauses, and by extension, of the attitudes, are propositions does not strictly follow from Fregean semantics and should be seen as an *ad hoc* stipulation with respect to that theory.

On the face of it, Russell's theory seems to fare somewhat better with respect of propositionalism, because it does not take definite descriptions to be referential terms. In this way, Russellian semantics avoids the implications from the slingshot in that it rejects theorems (D) and (F) in the argument.[101] However, for that very reason, Russellian semantics *does not* imply propositionalism since it does not construe the subordinate sentence as a singular term. Instead, Russell opted for a multiple-relation construal of attitude sentences which invoked as many relations between the agent of the attitude and the propositional constituents as their number combined. For example, the sentence 'Othello believes that Desdemona loves Cassio' involves a 4-place relation between Othello, Desdemona, Cassio, and the relation of loving. I think the problems with this proposal are well known and so I won't pursue it here. I will just note that it surely cannot supply a notion of content fine-grained enough for our psychological theorizing.

We can now see that in contrast to what is commonly thought, strictly speaking, propositionalism does not follow either from Frege's theory, nor from that of Russell, which leaves us a bit in the lurch. For propositionalism seemed so far our best shot at fine-graining the attitudes, given that we construe propositions either as made up of modes of presentations, as is the case according to Frege's theory, or as fine-grained truth-conditions, as follows from Russell's theory. Indeed, in the last chapter I have shown reservations with regard to both theories, but then, even my hybrid theory which combined Fregean and Russellian semantics required propositions as the objects of the attitudes.

At this point there seem to be two courses of action open for us. One is to reject the claim that the logical considerations invoked in section 2 of this chapter

---

[100]Of course, consonant with Frege's own semantics, we could say that the two sentences still differ in their expressed sense. But we remember that as per the relational theory, the issue is the reference of 'that'-clauses, not their sense.

[101]I have later learned that this argument was in fact already advanced by Gödel (1944).

are decisive in favor of the relational theory. That is, we can claim that notwithstanding such formal niceties, logic should not have the last say on material issues, and that in particular, there is no conclusive reason to suppose that the logical analysis of ascription sentences needs reveal anything substantive about the structure of the attitudes. In which case we can simply deny that intentional states involve a relation to *objects* such as states of affairs, facts, propositions, truth-conditions, etc. Instead, we can claim, in line with the discussion of Devitt's proposal, that intentional states (or even 'that'-clauses, if we want to preserve something from the logical analysis above), *express* representational properties, although I have already conceded that such representational properties might be just our familiar propositions in a new logical category.

Then there is the second course of action, that I think is in fact compatible with the first, which is to counter the slingshot head-on. For if we take a second look at the argument, we would notice there an assumption which should have raised our suspicion from the very get go: This is the first assumption according to which logically equivalent sentences express the same proposition. It seems that in any case we would have to attend to that assumption for surely, even if we had avoided the consequences of the slingshot by denying that 'that'-clauses are referential terms, we would have ended up with a conception of propositions which is too coarse given our empirical constraints (see in particular chapter Two, section 6.B). But that this assumption is implausible can be seen also from considerations outside the field of intentional psychology. This is because it has the absurd result that all the true sentences of logic and mathematics (given logicism), to the very least, express the same proposition.

Now Frege himself was ambivalent on the question of logically equivalent sentences. On the one hand, he did express himself to the effect that if two sentences are logically equivalent, then they express the same thought.[102] But then in the Grundlagen, Frege denies that all the truths of arithmetic express the same thought (in Frege 1964:35).[103] His worry was just the one mentioned in the last paragraph that mathematical propositions such as 2+2=2+2 would express

---

[102]In a letter to Husserl, dated December 9, 1906 (in Frege 1980:70). What Frege in effect says there is that two sentences A and B express the same thought if it is logically impossible for them to have opposite truth-values. For a more detailed analysis of this correspondence, see L. Weitzman 1997.

[103]In the above mentioned letter to Husserl, Frege indeed qualifies his aforementioned claim that the logical equivalence of two sentences is sufficient for an identity in their senses, by exempting logical truths from it. But then Frege offers no other criterion for identity of thoughts or logical truths.

the same proposition as 2+2=4, from which the puzzle again would arise as to how anyone can come to learn anything new by learning the second truth. But what made Frege even think that logical equivalence was sufficient for identity in sense? Well, we know that Frege accepted the principle that logically equivalent sentences have the same truth-conditions, in the sense in which they are true or false under the same overall conditions (what we call today possible worlds). Which means that no rational and competent speaker of the language could hold either true while not the other. But that is exactly how Frege saw the situation with respect to sentences which express the same thought (in Frege 1892b).

Now, it cannot be denied that there are some semi-Fregean approaches to propositions which endorse this very reductive conclusion, for example, the Carnapian possible world approach to semantics (Carnap 1947). Like Frege, Carnap assigns to every semantically relevant expression two semantic properties: an extension and an intension. Intensions are similar to senses except for the stipulation that they determine the extension of an expression not directly but relative to a possible world. In general, it is common to construe the intension of a singular term as a function from worlds to individuals, of predicates from worlds to sets of individuals, and of a sentence from worlds to truth-values. In this way, a proposition can be viewed as the set of all possible worlds in which the sentence is true. Clearly this has the result that all necessarily equivalent sentences, that is sentences which are true in the same possible worlds, would have the same intension, and thus express the same proposition.[104]

As we can see, Carnap's semantic theory should be seen as a reductive theory of propositions, construing propositions in terms of either sets or functions on worlds and truth-values. The approach is then nominalistic, where the constraints put on the ontology are different from the ones that a naturalistic theory of meaning aspires to respect. In any case, this approach is not one which friends of properties, I would argue, need to consider too seriously (see also my remarks on nominalism in note 94 above). But then, if we agree to reject this extensionalist approach to propositions, we can gain two valuable things: First, if logically equivalent sentences do not necessarily express the same proposition, then on whatever conception of representation that we eventually decide to adopt, it will help keep the grain issue in check (a desideratum that was established in my discussion in section 6.B of chapter Two). Second, we can now see a way to block the slingshot argument, with the benefit that 'that'-clauses can continue to play the role of singular terms as follows from Frege's semantics. This even if we now take the attitude relation to be a *functional* relation between

[104]For a contemporary example of someone who advocates a serious consideration of that view, see Stalnaker 1976, 1984:2.

the agent and the event of *instantiating* some representational property, as suggested above, rather than between the agent and a propositional *object* directly.

By rejecting the principle of logical equivalence, I allowed that necessarily equivalent sentences, and hence 'that'-clauses, can refer to or instantiate distinct propositions. But still, there is the issue of how to grain those propositions (or representational properties, as we now might call them) in a way which would respect our empirical constraints. That is, it is one thing to reject a certain individuation scheme for the attitudes, quite another to offer an alternative one. In chapter Three and in this one I implied that we need a fine-grained *truth-conditional* theory as a theory of propositions. In the next section I shall review one attempt to formulate such a theory, which is Bealer's theory. Were it to succeed, we could accept the Fregean analysis of attitude sentences with a 'that'-clause referring to a *Bealer proposition,* and derive from it an adequate conception of intentional states.

4. Bealer's Theory of Propositions

Perhaps the most courageous attempt to come up with a fine-grained theory of propositions is George Bealer's theory of Properties, Relations and Propositions (henceforth PRPs). At the heart of Bealer's approach is his objection to the nominalistic program encountered above, which attempts to reduce intensional entities to either extensional functions or sets. Consequently, Bealer advertises a non-reductionistic perspective which takes PRPs as basic irreducible entities. Then, following Plato's idea of the 'weaving together' of the Forms, Bealer advances his calculus of universals defined mainly over the basic Boolean operations on propositions (Bealer 1983, 1993a, 1993b, 1998).

Bealer's approach to propositions takes them to be essentially algebraic operations (which include conjunction, disjunction, negation, a predication relation, existential generalization, etc.,) on some set of basic, *sui generis,* entities. But since PRPs also have semantic properties (such as reference and truth) Bealer adds to the above an extensionalization function (H) which assigns an extension to each entity in its domain. Propositions are assigned truth-values, properties are assigned sets of objects, relations are assigned ordered *n*-tuples of objects, etc.[105] For example, the semantics of the conjunctive proposition p&q is given as follows: H(conj(p,q))= true iff H(p)= true and H(q)= true.

[105]I skip here certain complications concerning the independent existence of propositions from their constituents. For a discussion see Bealer 1998. All that matters for our discussion is that for Bealer, a proposition such as 'Hamlet is Hamlet' would exist even if Hamlet doesn't.

So far we see no significant differences between Bealer's semantics and that of Frege/Carnap. The differences come out mainly in the semantics of the predication relation. Here Bealer distinguishes between two kinds of predication relations, a singular predication relation and a descriptive predication relation. The singular predication relation is the standard logical operation which takes properties and items in the domain to give propositions. For example, the singular predication on *F* and *a* gives the proposition *that a is F,* a proposition which is true iff *a* is in the extension of F. But singular predication is not suitable to cases involving definite or indefinite descriptions. This is because a singular predication involved with e.g., the sentence 'The prime minister of England is in China', would be considered true iff the item corresponding to 'The prime minister of England' in the domain is in the extension of 'things which are in China'. But in my extended ontology, this would be the individual *concept* 'The prime minister of England' which is not what we want.

Because of such considerations, Bealer introduces the notion of *descriptive predication* in something akin to Frege's notion of second-level predicate (or predicable).[106] A descriptive predication would not treat descriptions as objects but as intensional entities which can be satisfied by objects. This would then ensure that the semantics of 'The prime minister of England ' would be given by reference to the extensionalization function on the definite description, rather then by reference to the concept itself. The truth-conditions which follow, then, are the customary truth-conditions (as in Tarskian semantics), where the sentence would be true just in case whatever unique object satisfies the description (if there is one) is also in the extension of the predicate.

Descriptive predication can also become useful for the semantics of sentences which involve indefinite descriptions as in 'Whoever shot Kennedy is crazy'.[107] According to S. Neale, 'Whoever' is a 'number-neutral descriptive operation (*whe*) which takes a property to give the number neutral property. For example, to give the number neutral descriptive property of *(whe*(Shot Kennedy)). Using the descriptive predication, we could then predicate the property of being crazy on this descriptive property in an analogous manner to the descriptive predication in the case of definite descriptions. The truth-conditions will be determined in the same way as above, which means that the

[106]Frege thought of a second-level predicate as an expression which is not a name but which can still form a proposition from a first order predicate. 'Someone' is such an expression, as it would give the proposition 'Someone loves Betty' by being predicated on the first level predicate '--loves betty' (Frege 1952:38).

[107]Taken from S. Neale 1990. See also Bealer 1998.

sentence is true iff the individuals satisfying '*(whe*(Shot Kennedy))', if there are any, are in the extension of the predicate.

As can be seen, by introducing the notion of a descriptive predication, Bealer is able to fine-grain his propositions more than their truth-conditions allow. But this should not mislead us into thinking that Bealer's propositions are a more sophisticated version of Fregean propositions, that is, those which include modes of presentations as constituents. Instead, Bealer sees his propositional theory to be an extension of Russell's theory in which propositions include as constituents objects and properties themselves. Indeed, according to Bealer, the only differences between propositions and what he calls "conditions" that is, the "weaving together" of qualities and connections ("plus perhaps subjects of singular predications") is in the kind of logical operation which is at work in each case: "the thought [i.e., proposition] is formed by means of thought-building operation of predication whereas the condition is formed by means of the condition-building operation of predication" (1982:188).

5. On Conception 1 and Conception 2 Entities

Even though Bealer takes thoughts to be built from the same basic constituents as conditions, there is a fundamental difference between them which goes beyond the claim that they involve two distinct kinds of predication relations. This difference corresponds to the distinction between two conceptions of the identity conditions on intensional entities: what Bealer calls conception 1 and conception 2[108].

According to conception 1, intensional entities are identical iff they are necessarily equivalent. For example, according to conception 1, the properties of being trilateral and being triangular are identical, and similarly the condition that something is a triangle and that it is a triangle and such that 5+7=12.[109]

---

[108]Bealer seems to ambiguate between a distinction between two conceptions of *identity conditions* and between two conceptions of *entities*. To cut down on unnecessary complications, let us just say that a conception type *entity* is just one whose identity conditions conform to that type.

[109]Bealer does not argue for the logical equivalence of such properties as *being trilateral* and *being triangular* but merely assumes it. It seems that this assumption is based on the necessary co-instantiation of these properties (see p. 3 in his 1982 for an example). Thus, and more generally, it seems that Bealer accepts the nominalist principle that for any two properties F and G, F and G are necessarily equivalent if(f?) they are necessarily co-instantiated. But this principle is clearly false: necessarily the properties of shape and color are co-instantiated,

According to conception 2, necessary equivalence is a necessary but not a sufficient condition for an identity of intensional entities. Here entities are identical iff they are necessarily equivalent and such that they satisfy the constraint that none figures in "a unique non-circular definition" (1982:2) of the other.

Take for example the so called "Cambridge" property of grue. It is defined as follows:

(7) x is grue iff x is green if examined before t and blue otherwise (henceforth green/t/blue).

According to conception 1, the properties of grue and (green/t/blue) are on a par since all that matters is that a necessary equivalence can be established between them. Grue, then, is just (green/t/blue). Similarly, according to conception 1, green can be defined by reference to

(8) x is green iff x is grue if examined before t and bleen otherwise,[110]

However, according to conception 2, only (7) can count as a correct definition (i.e., give identity conditions) but not (8), since green already figures in a unique definition of grue. Hence, given that (7) is *the* unique definition of grue, (8) would necessarily give a circular definition of green in terms of grue.

From the above it follows that according to conception 2, the property of green and grue/t/bleen are not the same property, and similarly the properties of being trilateral and being triangular, being trilateral and such that 7+5=12, and so on.[111]

---

but still distinct properties.

[110]Bleen is defined as blue if examined before t and green otherwise. All these examples are from Goodman 1955, pp. 75ff.

[111]We define 'triangular' in terms of 'a closed plane figure having three angles', and 'trilateral' in terms of 'a closed plane figure having three sides'. Both are their unique definitions. If in addition we will define 'trilateral' in terms of 'triangular' as in 'x is a trilateral iff x is a closed plane figure having three angles', it would follow from the uniqueness condition that the property of being an angle is identical to that of being a side. Otherwise the former definition would no longer be the unique one. But surely the two properties are not the same. Here it is not the problem of circularity which threatens this definition but only that of the false implication from the uniqueness condition. Ditto for any other coextension with 'being trilateral'.

On the basis of the distinction between these two conceptions, Bealer constructs his theory of propositions. While he takes qualities, connections, and conditions to conform to conception 1, he takes propositions to conform to conception 2. In which case, propositions would be more fine-grained than truth-conditions since they would not be reduced to necessary equivalences. Then, if we construct the attitudes as relations to such propositions, we would arrive at a conception of intentional content which is finer-grained than what we have seen so far either in Frege or in Russell's theory.

However there are still some gaps in this theory. For one, Bealer understands that as it stands, his theory would have the implausible consequence that such "properties" as grue and bleen would count as genuine conditions, since as we just saw, they conform to conception 1 entities. So, Bealer qualifies the definition of conception 1 entities by introducing a further distinction between simple and complex qualities. Only simple qualities are subsumed under conception 1, thus excluding grue and its ilk.

The picture that one gets from Bealer at this juncture is the following. The world is constructed out of conditions which are logical operations over simple qualitites and connections. These very same simple qualities and connections make up the basic conceptual constituents of propositions. Thus, basic concepts are just the qualities and connections which make up the structure of the world, although the operations which are involved in their construction are different from those involved in the construction of conditions (we are not told in what that difference consists). In any case, this means that simple concepts, such as *chews* and *masticates*, still conform to conception 1. Complex concepts, on the other hand, which are constructed from simple ones by the thought-building operations, conform to conception 2. Accordingly, *grue* becomes a conception 2 entity, a *concept*, just like the concept *being triangular*, *being triangular and such that 2+5=7*, etc., which means that any necessarily equivalent concepts to these are not thereby identical to them.

Now before I go on to discuss and criticize this theory, we can surely applaud the stipulation that logical equivalence is not sufficient for conceptual identity, as this can be used to explain some cases of failure of substitutivity, as in my examples of the propositions (p&q) and ~(~pv~q) (see also Bealer 1993b). Hence, if we could adopt this conception of content (conception 2), we would certainly be on our way to establishing the kind of empirically adequate intentional content we were looking for.

Still, there are some issues and questions that Bealer's theory needs to answer. The first concerns the distinction between basic and complex qualities. We saw that in the first formulation of the two conceptions of identity conditions, all that was required was that some properties be more basic than others in order to avoid the circularity in definitions. For example, I had stipulated that green was a more basic property than grue in order to avoid

defining grue in terms of green, and then green in terms of grue. And surely, some would wonder whether it should not be the other way around, which would mean taking grue as the more basic property, although perhaps all that this shows is that we cannot always know which properties are more basic than others. After all, as I understand Bealer's theory, it makes a metaphysical claim about the structure of conditions and concepts, not necessarily an epistemological one.

But be this as it may, the point that I want to get at here is that as per the first formulation, it was still open that green itself would be less basic than some other quality, and so definable in terms of it, and so on ad infinitum (as Bealer's parenthetical remark on p. 2 makes clear). But now on the latter formulation of identity conditions, where the distinction is made between complex and simple entities, one would have to come to bedrock at which certain properties are strictly basic, since otherwise all concepts/properties would become conception 2 entities. However, I would argue that it is hard to see how *any* concept or property can be said to be basic, by which we mean that any analysis would stop with it.

Again, let's take green as an example. Even on the supposition that we hit rock bottom when we attempt to supply a *conceptual* analysis of green, that is, in synonymous terms, that is surely not the case when we turn to empirical or scientific definitions. That is, we can define green by reference to light waves of some frequency or alternatively, by reference to paradigm samples ("x is green iff it bears the relation 'same color as *this*'"), etc. Would that mean then that green is on a par with grue (both being complex entities)? Hardly I should think.

Another problem with Bealer's theory is the implication that if two necessarily equivalent concepts differ in their cognitive value, then at least one would have to be an essentially complex concept. But this seems to me to be too close to the descriptive theory that was discussed and rejected in chapter Three, the theory according to which differences in cognitive value imply differences in the associated descriptive content. But in chapter Three 3 I said that whatever descriptive content is associated with a term cannot be *constitutive* of its content, since different and distinct descriptive contents can be as successful in securing its reference. That is, the point I argued for in chapter Three was that even if a complex description can be associated with a concept, this does not mean that the concept itself is complex. But this is just what follows from Bealer's theory.

Further still, if necessarily equivalent concepts can differ in their cognitive value only if one or both of them are complex, then what are we to do with such simple concepts which still seem to show differences in their cognitive value, for example, the concepts 'chews' and 'masticates', 'renate' and 'cordate', and so on? Of course, one might object here either that these are not simple concepts after all, following our strategy of giving scientific definitions, or alternatively that contrary to appearances, they do not differ in their cognitive values. In fact, it

seems that both claims go together since it is hard to see how there could be more than one true scientific definition for these pairs, which means that science would give both concepts the same cognitive content. In which case, we would have on our hands a reductio argument against the idea of giving scientific definitions to these concepts if we would have an independent reason to think that they *do* differ in their cognitive value. And I will later claim that we have one such reason and it is called the Mates' criterion. If so, then Bealer's theory cannot account for differences in the cognitive content of such concepts merely in virtue of their conceptual complexity.[112]

6. Naming trees, Causal Theories and Non-Platonic modes of Presentation

In light of the apparent difficulties of fine-graining meanings by reference to 'the things themselves', or by reference to modes of presentations, ever since Frege presented his puzzle, philosophers were attracted to some form of meta-linguistic solution, one that would solve all the apparent cognitive differences by reference to linguistic ignorance. In this sort of solution, which is also Fregean in spirit in that it relies on modes of presentation albeit of a special sort, ignorance of conventions about naming or naming practices has been substituted for ignorance of senses. After all, what could be more simple than ascribing the differences in informational content between, say, "Hesperus" and "Phosphorus", to their being different *names*, with different orthography and phonetics? For surely, if Jane assents to 'Phosphorous is a planet' but not to 'Hesperus is a planet', that is because she does not know the *meta-linguistic/semantic* truth that '"Hesperus" refers to Hesperus'. Or alternatively, she does not know that Phosphorus, or Venus, is also *called* "Hesperus". Similarly, it would seem that the easiest and most intuitive way is to suppose that the differences between, say, "chews" and "masticates", are lexical differences, such that do not learn anything *about the world* when we open the dictionary and learn that they are synonyms.

Probably the most famed theory that has taken this road is that of Carnap (1947, cf. Church 1954:66-7), where differences in cognitive values were mapped

[112]In Bealer 1993b, there is the attempt to account for Frege's puzzle and the failure of substitutivity of coextensive terms exhibited by the Mate's Criterion as another instance of a distinction by logical form. He says: "Frege's original puzzle about identity sentences is in a sense solved in this hyper-fine-grained setting. The reason is that the relevant 'that'-clauses - [that a=b] and [that a=a] have a different logical form" (p. 32). I have to admit, though, that I don't understand this theory as I can see no distinction in logical form between 'that a=a' and 'that a=b'.

onto differences in the linguistic forms that the attitudes were related to.[113] This simple (or simplistic?) solution to Frege's puzzle was in fact already anticipated by Frege in his discussion of the puzzle. He says there (1892b:56-7):

> If the sign 'a' is distinguished from the sign 'b' only as object (here, by means of its shape), not as sign (i.e., not by the manner in which it designates something), the cognitive value of a=a becomes essentially equal to that of a=b, provided a=b is true. A difference [in cognitive value] can arise only if the difference between the signs corresponds to a difference in the mode of presentation of that which is designated

Frege's point, as I see it, is that not every difference in a linguistic property can account for a difference in information value. Surely, some such differences do convey certain information about language use. It is certainly an acquisition of knowledge if I learn that Bill Clinton was called 'Blyth' in his youth, or that 'Rhodesia' is the former name of Zimbabwe. But this is knowledge of conventions and not what we learn from true identity sentences, or at least not always.[114]

Indeed, it seems to me that what might be confusing about this issue is that sometimes the knowledge that identity sentences convey is just such knowledge of linguistic conventions. For example, suppose that without engaging much in my friends' discussion, I overhear them talking about some Bill Blyth. I ask them who it is that they are talking about, and they reply 'Bill Blyth is Bill Clinton'. This would more probably be a case of conveying linguistic information, if there is still no other information that can I associate with 'Bill

---

[113]I will not get here into the question of what 'form' means in this context. Carnap took his theory to be a syntactical theory, just like Fodor's RTM, but with the difference that his sentences were in a public language, whereas Fodor's are sentences of mentalese. But to be sure, syntactical form is not the same as morphology, since two predicates can have the same syntactical form even if they differ in their shape (see also Devitt 1996:ch. 5).

[114]Of course, any *a priori* reliance on the distinction between knowledge of linguistic conventions and knowledge of facts to determine what belongs in content and what does not would be guilty of presupposing the analytic/synthetic distinction. But one can adopt an *empirical* method to make for that distinction, as I will suggest later in the chapter. In which case the differential contributions of world and linguistic conventions do not enter into the *constitutive* conditions on content (barring further assumptions), and hence do not commit the fallacy of the A/S distinction.

Blyth'. But suppose that the situation is somewhat different and that I ask this question at the end of the discussion, after I learned quite a lot about the individual Blyth. If I'm then told that Blyth is Clinton, the information I acquire is not merely linguistic (although that too), but about the world. I then 'extend my knowledge' about the world since I can now make predications on Clinton/Blyth which I could not make before.

Neo-Carnapians (as I will call them here) understand that linguistico-morphological information on its own cannot solve Frege's puzzle. They therefore suggest something more sophisticated in the form of naming-trees (Bealer 1998)[115] or d-chains (Devitt, 1996).

According to Bealer (1998), the solution to Frege's puzzle lies in the introduction of what he calls "non-Platonic modes of presentation" (in Bealer 1998). A non-Platonic mode of presentation is a mode of presentation which is symbolically, rather than property, based. That is, a non-Platonic mode of presentation presents by way of constructed entities - linguistic structures which extend in time and space. For example, Bealer argues that the way names provide their users with access to their nominatum depends on the particular practice of their use, a practice which can extend to historical as well as contemporaneous occasions of their use. The idea is that the use of two different names to designate the same object might correspond to two differing naming practices, or "historical naming trees". I might be able to shed some light on Bealer's conception by considering a similar position advanced by Devitt, only that in the latter's theory "historical naming trees" makes way for "d(esignating chains)":

> [A d-chain] starts with a person's face-to-face perception of an object, a "grounding", and may run through many people by the device of "reference borrowing" in communication. There are usually several d-chains involving the one object and word, all linked together to form a network...I identif[y] the meaning of a name with the property of referring by a certain type of d-chain, each name by a different type." (1996:164-5)

Now, before I continue to discuss this type of a theory, I would like to digress a bit and mention a few distinctions that cross-classify with the various sociohistorical semantic theories that one can find in the literature. To begin, one thing that historical-chain proposals seem in general to imply is some sort of

[115]As we shall see below, Bealer 1998 is a quite different Bealer than Bealer 1982. In fact, as I see it and present it here, by going meta-linguistic, the later Bealer departed quite a long way from the intensionalist solution I examined in the last two sections.

causal relation to the object that makes for the "grounding" of the chain, and perhaps also in the way it "spreads" in time and space (Devitt ibid., but most famously, Kripke 1972, and Putnam 1975). Of course, since much of this chain is intentional, one risks certain circularity if the idea would be to ground differences in intentional content in differences in d-chains or naming trees. I will return to that problem later but in this context it should be emphasized that not all historical theories of meaning are *reductively* causal, in that they pledge to translate or substitute all semantic terms with causal terms. Some are 'merely' causal, in that some causal relation plays a definitive role in the theory (perhaps in the 'face-to-face perception of an object').[116] In which case, the bulk of the work in explaining cognitive difference will have to be done at the intentional level, which is just the phenomenon we are trying to explain.

In light of this, it seems that, usually, when philosophers advocate a causal theory of meaning, it is a direct reference theory they have in mind, not necessarily a reductive theory. Sometimes such scholars object that their theory is not a causal theory of *meaning* but of *reference* (Devitt, ibid., p. 170), anticipating the objection that two terms caused by the same object would have thereby the same meaning (which would leave the solution to Frege's puzzle to be worked out at the level of sense). However, such a claim would have any force only if the theory of meaning is not itself a theory of reference, i.e., is not a 'Fido'-Fido theory. On the other hand, there are those who have claimed that not every 'Fido'-Fido theory of meaning would fail to solve Frege's puzzle. Thus, McDowell has argued that it is possible to construct a referential theory of meaning in which certain co-referring terms would exhibit cognitive differences (in McDowell 1977). That is, McDowell thinks that while, e.g., "Hesperus" and "Phosphorus" can be said to have the same meaning (reference), this is still compatible with '"Hesperus" refers to Hesperus' not expressing the same *truth* as '"Phosphorus" refers to Phosphorus', in the context of a Davidsonian *theory of meaning* (ibid., p. 169-170).[117] But again, due to the essential intentional

---

[116]A similar point is made by Almog (1984:488) who claims that even Kripke's theory cannot be taken to be *a causal* theory since Kripke never intended it to be *reductive*.

[117]As I briefly mentioned in chapter Two and in this one, the idea behind Davidson's theory of meaning is to use Tarski's convention T sentences ('"s" is true in L iff p') as a basic format for a theory of meaning. In effect, Davidson capitalizes on a remark by Tarski (in Tarski 1944:350) that the used sentence 'p' can be seen as a translation of the mentioned sentence (thus avoiding the need that the meta-language *include* the object-language). Because of this Davidson sees his theory as 'an inversion' of Tarski's theory, assuming truth as basic to

nature of such theories, such a solution cannot account for the cognitive differences between intentional states. That is, from my perspective, the task is to explain how such 'truths' can express cognitive differences, and then use this in turn to explain how people can respond differently to them, whereas a Davidsonian theory of meaning merely takes such differences for granted.

In this I end the digression and will now return to the meta-linguistic solution of Bealer and Devitt. According to Bealer's theory of propositions as involving non-Platonic modes of presentations, a proposition such as 'Tully is emulated' would be represented as a descriptive predication of 'being emulated' on the non-Platonic mode of presentation corresponding to the proper name Tully thus: $Pred_d$(being emulated, "Tully"). On the other hand, the proposition that Cicero is emulated would be represented as $Pred_d$(being emulated, "Cicero"), and so on. Then, we might assume, the same picture could be extended if the proposition would be that Tully masticates, only that this time both modes of presentation are non-Platonic: $Pred_d$("Masticating","Tully").

Bealer thinks that his theory can also account for Kripke's puzzle about belief (Kripke 1979). That puzzle concerns Pierre, a Frenchman living in France, who only hears good things about London. Hence he forms the belief that he would express by "Londres est jolie". Pierre then moves to London and after some harsh experiences with his surroundings, forms the belief he expresses in English as "London is not pretty". Using a translation principle, according to which co-referring names in different languages express the same meaning, and a disquotation principles, according to which the sentences people express or assent to express their beliefs, Kripke concludes that Pierre both believes that London is pretty and that London is not pretty. Since intentional realists think that these different beliefs can systematically give rise to different behaviors, they would have to say that they express different contents.

---

derive translation or sameness of meaning.

Now as both Tarski and Davidson accept that the construction of T sentences should follow basic compositional canons, a semantic theory which assigns semantic properties to complex expressions on the basis of those of their sentential constituents has to be first developed. In fact, at least in the case of a Davidsonian theory of meaning for a language, the construction of such a theory is constrained by the linguistic responses of the agents for whose language the theory of meaning is intended as a theory of interpretation. But this implies that there might be more than one way to derive a T sentences for which there is a consistent assent from either the agent or the linguistic community. As far as I understand McDowell's point, those differing ways of getting at the same extensional truth count as different 'truths', hence his solution to Frege's puzzle.

It is interesting to note that this puzzle is not just an embarrassment to descriptivists who take differences in cognitive values to imply differences in descriptive contents. It is also an embarrassment for the direct reference theory since both 'London' and 'Londres' refer to the same object.

The way Bealer thinks he can account for the differences in cognitive content of the two sentences (and hence of the two beliefs) is by reference to what he sees as their distinct non-Platonic modes of presentation: $pred_d$(being pretty, "Londres") in one case, $pred_d$(Negation(being pretty, "London")) in the other case. And we might expect Devitt's solution to be the same, that is, that the two names connect to two different d-chains.

However, here things begin to get complicated. One wonders what is the exact relation that such linguistic chains bear to the content of the intentional state (or expression) they attach to. Surely, people cannot be expected to know everything or even most things that are going on in such chains. Alternatively, it is possible that most of what belongs in the chains for two names would overlap but still there are differences in its remote corners which might affect significant cognitive differences. Would this matter to the content of my thought and if so, to what extent? Does this picture imply some sort of resemblance theory of cognitive content which would then presuppose identity conditions for "networks" for names? Indeed, Devitt thinks that such identity conditions need to be given by reference to the physical type the token name belongs to (*pace* Frege's objection) "linked together by ...[sorts] of mental processing in the speech community" (ibid., p. 168). I hope I need not argue that this seems quite mysterious if not downright circular.

Perhaps due to considerations such as these, Bealer cautions us not to interpret his meta-linguistic theory as giving the *meaning* that is conveyed by the proposition, since it is not required that an agent know the historical naming tree, or the naming practice, in order to understand it. Rather, the idea is that the meta-linguistic structure of non-Platonic modes of presentation is shown only in their logical analysis, which he takes to be non-identical to their conceptual analysis. The consequence of that is that such meta-linguistic analyses would figure in attitude reports only as external indexes. In other words, they would not purport to underwrite a commitment to what the content of the target mental state actually is, in the sense of what the agent can be said to know or understand(implicitly or explicitly).

Now, while Bealer calls this result or stipulation of his theory "interpretive liberalism", it seems to be quite out of line with Bealer's own realism about mental states. After all, as intentional realists, it would seem that our interests in such cognitive differences should be, at least in part, in their differential contribution to the causal powers of mental states. And I have followed Bealer so far because I thought that this is the kind of content I would be getting from him. But once Bealer concedes that such differences are merely "external" (an

idea we shall discuss in much more detail in the next chapter), then it is hard to see how they could be translated into causal efficacy, and therefore, how this theory could be of any help to us.

Let us go back to the question raised above, in what would consist the contribution of naming trees or designation chains to the differential cognitive values of two coextensive terms. Well, we say that it cannot consist in the nominatum (object or property) since this one is assumed to be the same for both names. Hence it must be in something else that two distinct chains or trees differ. What could that be? One way to think about it is that the chains differ in the thoughts and intentions that users associate with the expressions. For example, the thoughts about "hesperus" that it refers to Hesperus, or that it refers to whatever people in my linguistic community refer by it, etc. Well, if the content of these thoughts are taken in their tokenings, we would get involved in a circularity. For as I said, it is just the content of this token thoughts we want our theory to explain.[118] Alternatively, if the content of these states is taken in abstraction from the occasions of their instantiation, we would just get Fregean senses in the back door. For contents, or meanings, which are independent of their particular tokenings are Platonic in exactly the same way that Fregean senses are assumed to be. Of course, advocates of such a theory might not think that these abstract conceptions *determine* reference (whether Frege ever did or not), only that it would supply the cognitive content required to account for Frege's puzzle. But as was my point in chapter Three (mainly sections 2 and 3), whatever plays the role of cognitive content has to be veridical in order for it to be informational. And then we are back to the objection that the having of true beliefs cannot be a condition on content (even if it has nothing to do with fixing a reference). That is, if the cognitive content of our terms/thoughts depends on the true conceptions we associate with them, then the having of *false* conceptions (as that the people around me refer to Socrates the philosopher when they use "Socrates", although in fact they refer to the Brazilian soccer player), will stand in the way of successful reference. In any case, it is hard to see how that theory can account for the information content that we acquire from true identity sentences concerning the *object* of reference, which is, after all, what intuitively seems to be the case. Do we really believe that when we learn that the morning star is the evening star (except in a situation of the type I alluded to above in the 'Blyth' example), we learn facts about our linguistic community and the linguistic

[118]Bealer (1998, n. 35) in fact alludes to such a circularity but claims that it can be solved. But as no solution for this is anywhere offered by him, and as I cannot see how such a solution *could* be offered, I suggest we just ignore this claim.

intentions associated with the use of such terms?[119]

More and more, I would argue, the meta-linguistic theory that Bealer and Devitt advocate begins to resemble some extended theory of conceptual role semantics, spread to include the historical and contemporaneous linguistic community of the agent. In which case, in addition to problems I have so far uncovered, it would inherit all the old problems I unraveled in chapter Two about compositionality, holism, and meaning incomparability. Altogether, I think I have established a sufficient ground for rejecting this theory.

7. On Propositions and Logical Closure

So far in this chapter I reviewed in some detail the theory according to which the attitudes are construed as relations between organisms and propositions. Two conceptions of propositions were the focus of the discussion: The Fregean conception of propositions as modes of presentations (by which I also included 'linguistic modes of presentation'), and the Russellian conception of propositions, exhibited in Bealer's theory, of propositions constructed from properties and relations themselves. I think the discussion up to this point has made it clear that neither theory is able to supply a notion of intentional content that can satisfy our empirical aspirations. How so?

Well, it seems to me that one of the problems in giving an adequate propositionalist theory of the attitudes lies in the fact that no satisfactory answer has been given so far to the question, what actually *are* propositions? We saw that propositions were taken by Frege and others to be the meaning of sentences, modes of presentations, semantically evaluable entities (that which is either true or false), the semantic values of 'that'-clauses, the objects of the attitudes, abstract entities, and more. But importantly, these characterizations only tell us what *roles* propositions might occupy, not what it is that occupies that role. We saw various attempts at giving *reductive* explanations of propositions, in terms of functions from possible-worlds to truth-values, or in terms of ordered sets of individuals and properties/relations. But still, the question remains as to what

[119]In fact, I even doubt that facts about other people's intentions play *any* role in *what* we learn when we understand a sentence. What I have in mind here is Grice's claims that to understand the meaning of a sentence p is to come to know a) that you, the speaker, want me to come to posses the belief that p (in Grice 1957) or alternatively b) that you believe that p (in Grice 1969). Of course, it is not difficult to find counterexamples to both conditions. In my view, to understand a sentence is to come to know strictly its truth-conditions.

makes these functions and sets in particular be *propositions,* rather then more generally abstract entities? After all, as Frege taught us, it is one thing to give a definition for something, quite another thing to justify it. So my question is, what justifies these reductive attempts of propositions? In fact, I doubt any of the theories I have reviewed can answer that question.

So to try to make some progress here, I would like to suggest a different kind of definition for propositions which is *not* reductive. My suggestion is that we define propositions as that kind of entity for which a certain set of logical operations are well-defined. More specifically, the claim is that for any entity x in our total ontology, x is a proposition iff L(∝,x), has a value for some ∝: where L is a well defined logical operation for ∝, and ∝ is a term (in the Russellian sense of terms) standing for an object, a property or a relation.

For example, suppose that ∝ is the collection <Socrates, being wise>, and L is the operation of singular predication. In which case L(∝,x) would give the proposition that Socrates is wise. Similarly, if ∝ is the collection <Socrates, Bachelor>, and L is the operation of logical entailment, then the result would be the proposition that Socrates is an unmarried male, the proposition that someone is a bachelor, etc. If L is material implication, then the value of L(∝,x) on <Socrates, Bachelor> would include the proposition that Socrates is a bachelor and 2+2=4, that Socrates is a bachelor and 2+5=7 and so.[120]

As can be seen, this definition deviates from other conceptions of propositions in that it makes no reference to semantic evaluability, although it can accommodate one, and the same goes for modes of presentation. The idea is that once one has defined what it is for a structure to be a proposition, one can proceed to the uses that such a structure allows. In particular, it can be seen that due to the logical behavior of such structures, they can be used as compositional modes of presentations, as compositional objects of thought, and so on. After all, almost anything can be semantically (e)valuable, but propositions are especially suitable since, *ex hypothesi,* they also have logical composition.[121]

---

[120]I assume here that the standard logical operations are all well defined independently. The purpose here is not to give the general foundations to logic, but to define propositions. The idea then is that given a bunch of atomic properties and a set of logical operations defined over these, we could derive propositions. This is not quite unlike Bealer's derivation of thoughts from qualitites and conditions with the difference that we do not recognize similar operations (what he calls "condition building" operations) that issue in *conditions.*

[121]Here someone might ask 'what about sentences?' as sentences are also closed under syntactical operations such as derivation (Fa & Gb entails Fa v Gb,

In any case, my purpose here is not to propose a full-fledged theory of propositions, but to show why I think propositions would necessarily fail to supply a viable notion of intentional content. If propositions are defined as structures closed under logical operations, and if intentional states are defined, as is the case with propositionalism, as relations to propositions, then the same logical closure of propositions would be inherited by the psychological. But the lessons we learned from Hume's stricture (in chapter One) and from the failures of substitutions in intentional contexts (in chapters 2-4), is that the psychological is not, and cannot be, closed under the logical. And indeed, it seems that what the various puzzles which I examined so far point to is that there should be a demarcation between the conceptual and the cognitive, even if there are many areas in which the two overlap. This reluctance of the cognitive, whether in the form of mental content or even linguistic meaning to "mirror" the conceptual, shows that an alternative notion of intentional content is called for. Propositions are ill-suited as intentional contents because, to repeat, by definition they are closed under logical operations. In which case, any attempt to distinguish two of them which are logically related will have to rely on questions of logical form or complexity, as we saw in Bealer's theory. But as with syntactic simples, conceptual simples do not have a logical form (or rather, they all have the same form); hence Bealer's attempt to rely on "Non-Platonic modes of presentations".

These considerations should make us conclude, I want to argue, that we were looking for the appropriate candidate for intentional content in the wrong place. We thought that we could somehow fine-grain intentional content if we turned to propositions as the objects of the attitudes, but we now see that propositions are in fact not fine-grained enough, even if they are Fregean propositions involving modes of presentations, or Russellian propositions involving hyper-fine-grained truth-conditions. For, after all the dust is settled about these theories (and there is much of it), we are still left with a notion of content which is not as empirically adequate as we require.

In my opinion, the only solution left for us is to go informational 'to the things themselves', but to the things as finely-grained as mental content happens to be.[122] *How* finely grained, we shall see in the next section that also ends this

---

etc.). I have no objection that in this case we would regard sentences as propositions, provided they are *defined over* these derivations. But wouldn't it be then just simpler to say that sentences *express* the propositions defined over these operations? The only complication I see is that we would have to add syntactical objects to our ontology but I see no particular problem with that.

[122]This would exclude, by the way, the sort of 'situation semantics' advocated by Perry and Barwise (1983). That is since while trying to make for the

chapter.

8. Mates' Criterion and the Method of 'Psychoanalytic Ontology'

It seems to me that the first step on the way to formulate the right theory of content for mental states is to reject the intuitive and a priori criteria that we have so far encountered for when terms or thoughts differ in their cognitive content. In chapter Three we saw that Frege attempted to define the notion of cognitive value by reference to two a priori criteria. According to the first, two terms have the same cognitive value if they can be substituted for one another in an analytic sentence without turning it into a synthetic sentence. The second criterion was that two terms have the same cognitive value if the substitution of one for the other didn't turn an a priorily known sentence into an posteriori one. Now, I have claimed that both criteria were untenable, given my acceptance of the no analytic/synthetic distinction. But in effect, in that I have rejected two a priori criteria for identity/difference in cognitive value on the basis of another a priori consideration.

First to the Frege criteria. Here Frege assumes, presumably by mere reflection on the concept of cognitive content, that ideally rational and competent speaker of the language could come to know whether a sentence was analytic, or known a priori. I should note that the same is the case with respect to Russell's theory which uses the principle of acquaintance to assign content. Here again it is assumed that we (i.e., Russell) could predict how every rational agent would decide in questions of cognitive difference, by being ourselves acquainted with the concept of cognitive content. Clearly, neither theory leaves much room for empirical research into the nature of cognitive content, and the conditions under which rational agents actually discern differences in cognitive value.

But then also note Quine's 'refutation' of the idea that people can, just by reflection, note whether two terms express the same concepts or not, or alternatively, whether two concepts are one and the same or different. In short form, his claim is that such an ability requires one to be in possession of a *definition* for analyticity (or sameness of meaning), whereas he attempts to show, again by a priori and conceptual means, that no such definition could be produced. But surely, this very formulation seems to go against the conclusion of Quine's own argument, i.e., the thesis of confirmation holism. According to confirmation holism, what confirms what is an a posteriori affair, which depends

---

distinction between some coextensive contents (for example, between "Socrates is wise" and "Socrates is wise and 5+7=12"), Perry and Barwise make no attempt to distinguish water thoughts from H20 thoughts etc. For them, both of these contents correspond to the same situation.

not on linguistic conventions but on the relations of properties in the world. But that means that even the a/s distinction would itself have to be subjected to the tribunal of experience, where we might find that some properties are *essentially* connected to others whereas others are only contingently so if at all. In other words, it seems to follow from confirmation holism that what relations in fact obtain between properties is a worldly matter which is possible for us to discover, given our psychological and intellectual capacities. In which case, we might learn from experience that the state of reality is such that there *are* some strong connections between certain properties which correspond to what we mean by 'analytic relations' and others which correspond to 'synthetic relations', and so on. This means that we might not only confirm in this case, a posteriori, the a/s distinction, but discover at the same time specific *essential* properties of certain properties. But what has all that to do with the notion of cognitive content?

Well, so far, in a few places in this thesis, I claimed that we require a notion of mental content which is not only semantically intuitive, but also empirically adequate such that it can pass the test of the empirical adequacy of theories and explanatory schemes which rely on it. The idea was that identities and differences in mental content, and hence in intentional states, must be decided by reference to the actual cognitive role of such entities in the psychological economy of rational agents, but not constituted by the latter. I assumed in addition that although the place where mental contents naturally show their powers was in the production and control of mental and behavioral processes, the origin of those powers lies *outside* of the mind, in the relations between the corresponding worldly properties themselves. The point is, though, that many, if not most, of those relations that minds can discern (for example, between the concepts 'analytic' and 'synthetic'), can be detected only in the way those properties *affect* minds, and hence "in" minds. This is where what I would call 'rational psychology' (somewhat after Fodor 1980) enters the picture.

If semantic relations between concepts are in fact the manifestation of property relations as they are recruited into the service of the production and control of behavior, and if this is their only manifestation which we (philosophers, scientists) can detect, then the way to map this intensional nexus is, I would claim, by doing a piece of '*rational psychoanalysis*' on ideally situated rational agents. To show where I'm heading, I would like, as a first approximation, to propose a way to connect this empirical notion of content with its semantic significance via the following semantic criterion for synonymy which was offered by Benson Mates (1951). I will call it the Mates' criterion (M) which goes like this:

M: Two expressions are synonymous [and hence have the same meaning] in a language L if and only if they may be interchanged in each sentence in L without altering the

truth-value of that sentence.[123]

The reason why I have singled out Mates' criterion for discussion in this context will be better appreciated, I presume, once we see that it allows us to individuate content by reference to *psychological* sentences of the sort that will become of interest to our rational psychologist. And an important implication of that is that we will now have a tool on our hands which would carve meanings much more finely than we have seen so far in any of the semantic theories I reviewed. For example, while Frege would treat pairs of expressions such as 'bachelor/unmarried male', 'masticates/chews','renate/cordate', as synonymous and hence as expressing the same sense, it is an open empirical possibility, given that L is the language of our psychological theory, that the Mate's criterion will treat them as distinct. Compare the two attitude sentences:

(9) S does not doubt that whatever chews chews,

and

(10) S does doubt that whatever chews masticates.

Let's us suppose that we find an agent who satisfies both (9) and (10). Let's also suppose that we make sure that S is a competent speaker of the language in which the sentences of the theory are coached (here English), that she is rational, and as much free from environmental influence and prejudice as we consider possible and relevant. What should we then conclude from the fact that S is willing seriously to entertain the possibility that a thing can chew without masticating? Certainly, our rational psychologist (and philosopher) should be interested to know, and here they might have a few options. The first one is that S, although a competent speaker, still she is ignorant of the relevant linguistic convention, but this possibility can be figured out and then eliminated. Another

[123]Although synonymy implies sameness of meaning, sameness of meaning does not imply synonymy. Unlike sameness of meaning, synonymy is a property of words and expressions *from the same language.* From this latter fact we can also reason why the following two sentences A and B are not synonymous, although they seem to satisfy MC: A. '"Banstickle" is Tal's favorite word' and B. 'the first word written on p. 659 of the O.E.D., is Tal's favorite word'. The reason the two expressions cannot be synonymous is that sentences which *mention* expressions *of* L are not *in* L. In any case, if L is the language in which our rational psychology is formulated, then mentioned terms of L are certainly not in L (in the same way that "Chemistry" is not a term in Chemistry).

possibility, which I will return to later on, is that the mechanisms through which S connects to the correspondent/s of 'chews' and 'masticates' in the world are distinct, which makes S oblivious to their presumed identity. But there is another option, and this is that the two properties are just distinct. This would be just a case where S serves as a good and reliable indicator of the fact that 'chews' and 'masticates' express distinct properties.

Here is another case. Consider the following three sentences:

(11) S knows that whatever is a circle is a circle.

(12) S does not know that whatever is a circle is a locus of points in the same plane equidistant from some common point.

(13) Whatever is a circle is a locus of points in the same plane equidistant from some common point.

Sentences (11) and (12) are attitude sentences, whereas (13) putatively expresses a definition. But as can be seen, these three sentences together create a puzzle (called the paradox of analysis, due to Moore, 1952). The puzzle arises since if (13) is true, then it follows that the property (concept) of *being a circle just is* the property of *being a locus of points in the same plane equidistant from some common point.* But then, how can S not know this fact, and if she does, what is the point in giving the analysis?

Well, by turning to MC, it can be claimed that the failure of substitution of the two concepts 'a circle' and 'a locus of points in the same plane equidistant from some common point' between the two psychological sentences (11) and (12) might just show, given the elimination of the other possible factors, that the concepts are in fact not one and the same.[124] This result is the application of the

[124]This theory is in some details close to that of D.F. Ackerman (1986). But while Ackerman's focus is specifically on the subject of the paradox of analysis, or more accurately, paradoxes of analysis (she would take sentences 9& 10 to pose a different problem than 11-13), she also thinks that the attitudes people form towards concepts can be revealing with respect to conceptual relations. However, there is perhaps a terminological dispute between us as Ackerman thinks that her method is a priori, apparently because it attends to people's rational intuitions concerning the concepts involved, and concerning counterexamples to suggested definitions. But surely, those 'intuitions', or introspections, as they are also called, should not be perceived as independent of the agent's experiences, if by that we mean the interaction of the agent with the instantiation of properties in her environment.

principle I now adopt that psychological sentences concerning people's attitudes are 'prior in the order of explanation' to the semantic facts. The psychological, in the sense of what ideally rational people would be willing to judge with respect to conceptual relations, takes dominance over the stipulative, the conventional, or just the semantic.

Now, this way of deciding general semantic questions by reference to psychological sentences and theories might seem to some to be foreign to Fregean and even Russellian semantics, with their disdain for psychologism. But, of course, there is no psychologism in this theory. This is because the use I make here of our psychological theorizing to carve contents is only to serve as a kind of litmus test for a *distinction/identity* in content, nothing more.[125] For to repeat, my assumption is that the origin of content lies in the informational (truth-conditional) relations that psychological states bear to properties instantiated in the world, not in the psychology of people as such. As McDowell says: "psychological explanations of behavior are central in the conception of a theory of language [i.e., meaning]...But their purpose is to confirm the descriptive adequacy of a theory, not to put an explanatory mechanism through its paces" (1977:168). I think this pretty much articulates my point here.

So far I have treated the 'psychoanalytic method' as abstracting from the linguistic, social, and natural environment in which agents are embedded, in a sort of methodological solipsism. Of course, this idealization is hard to approximate. In the case of language at least, as a social phenomenon (although *pace* Kripkenstein[126], not essentially so), sometimes the mapping of the intensional nexus (i.e., the relation of properties) cross-classifies with the linguistic conventions. This is because linguistic conventions are often a function of how much a certain piece of knowledge is wide-spread in a community. For example, a community in which most people would assent to (12) for appropriate substitutions of S, is one in which "circle" and "a locus of points in the same plane equidistant from some common point" *are* synonymous terms (e.g., the mathematical community). Still, this does not and should not mean that the two terms express the same *concept*.

I hope this example is sufficient to show the difference between my theory and one, like that of Burge (in Burge 1979, 1986), where the role that the linguistic community plays in determining the content of a term or thought is constitutive (I shall discuss this theory more fully in the next chapter). For me,

---

[125]In fact, and as we shall see in chapter Six, more a distinction than an identity.

[126]I.e., Kripke's interpretation of Wittgenstein on Rule following, in Kripke 1982. On Wittgenstein's argument, see his 1953.

the linguistic (and non-linguistic) conventions adopted by a community are a pragmatic issue which for the rational theorist form part of the 'noise' that has to be screened out in order to get at the heart of the matter, i.e., the relations of concepts. To repeat, what determines the content of linguistic terms in my theory are the reliable *mind-world* relations that the corresponding mental states bear to the instantiations of properties in the world. The Mates' criterion supplies only a *fallible* indication of what sort of relations these might be, one which has to take into account that sociolinguistic factors might not only serve towards getting at those relations, but at times drown them in the seas of language and culture. To sum then, my theory is that particular questions of the sameness or distinctness of contents are empirical questions, in fact doubly so. First, they are empirical questions since concepts/properties are real existents with real relations to each other which can be discovered through an empirical investigation. Second, they are empirical questions in that our best if not only way to track those worldly relations is through their role in human (and possibly non-human) psychological economy. The method of investigation into the nature of properties through their role in human psychology I called rational psychology, which can be conducted by reference to the rational verbal and non-verbal behavior of what approximates ideally situated rational agents. I have implied that the task here is formidable, as 'noise' coming from linguistic and other sources can easily mask real distinctions between concepts. In fact, I think it has masked a whole tradition in philosophy in the 20th century.

In any case, my concern in this chapter was not in how we can develop a method to discern concepts, but that we should be able to account for those cases where it is really distinct properties that make for the cognitive differences. All and all, and given what our rational psychologist would conclude, it might be that Hesperus *is* phosphorus, or that water *is* H2O, but it might be that they are not. In the latter case, the differences in the cognitive contents of the corresponding terms would be real, and we need to show what makes for that. For that I will have to develop a theory of properties which would be sophisticated enough to account for these differences without violating too many of our modal intuitions. I will attempt to do that in chapters Six and Seven. In the next chapter, I will return to the psychological question I left off above, and specifically to the issue of how to reconcile the wide individuation scheme of the attitudes that informational semantics brings with it, with the idea that the causal powers of mental states are local. This is the issue of individualism.

Chapter Five

Individualism

1. A Non-Intentional Core

So far in this thesis I have followed the intentional realist credo that mental states, as states of individual agents, are causally efficacious states which play a proprietary causal role in the production and control of behavior, as well as of other mental states. In chapter One we encountered a formal expression of this idea in the thesis about the nomological-intentional nature of PEPs: the idea that such practices involve the derivation of psychological explananda from "hedged" intentional laws. When I discussed this thesis I said that by the 'nomological-intentional' proponents of RTM mean *causal* intentional laws. Thus evolved the conception that mentalistic explanations were a species of *causal* explanations and predictions, constrained by the semantic relations between the propositional objects of the attitudes. I called the semantic constraint on the causal behavior of the attitudes, the isomorphism thesis (IT).

In chapter Two I went one step further and discussed at some length the functionalist view according to which to be a mental states just *is* to fill the causal role definitive of that kind of a state.[127] Of course, one does not have to adopt the nomological-intentional view of PEPs to be a functionalist. Indeed, we saw that this view about PEPs, that they are a species of causal explanations and predictions, was also shared by some of those who repudiated that view, either because they do not think that there are causal laws which subsume the mental *qua* mental (or rather, no *strict* laws), or because they do not think that one needs a law at all to ground a causal explanation. Thus, in the first group we have someone like Davidson who thinks that the 'because' in rationalizations involves a causal relation: not because it is grounded in mental laws but because it is grounded in strict physical laws (see chapter One section 6). Still, Davidson not only advocates an intentional-realist position about mental states or events - that they participate in causal interactions - but even a functionalist position regarding them, following from his functionalist position regarding events generally. For according to Davidson (1969b:179), events are identical iff they cause and are caused by the same events. This in effect means that Davidson's functionalism concerns *token* mental events, *defined* by reference to their causal liaisons.

Davidson thinks that the existence of a strict law is required to ground

[127] As Jackson and Pettit nicely put it, "the typology of mental states is a typology of causal roles" (1990).

mental causal explanations since causal explanations in general are 'Humean', presupposing generalizations.[128] Then there are those who repudiate the nomic-intentional thesis of PEPs not because they think that there is something wrong with mental or intentional laws, but because they think that causal explanations in general do not require laws. Here we can find someone like Baker (1995), who thinks that the right format for evaluating mentalistic explanations is through counterfactuals rather than by deriving them from strict or even *ceteris paribus* laws. And of course, there are all sorts of intermediate positions which although interesting, I will not pursue here.[129]

The battlefield over the question of the nomic character of mentalistic explanations that engages intentional realists, a dispute which at times spills over into the nature of causal explanations more generally, tends to conceal a deeper controversy concerning the metaphysical grounding of the relation of causation itself. That controversy is over the question of what has to be the case for mentalistic causal explanations to be *true*. In other words, the question is about the truth-conditions of mentalistic explanations so that these explanations would count as a species of causal explanations. The traditional positivistic answer is that mental properties and relations, as 'higher-level' entities, need to be reduced to properties and relations at lower-levels, eventually to basic physical properties and relations (Oppenheim and Putnam 1958; Smart, 1959). Ignoring here the issue of what is a basic property or relation, the idea is that for every property or relation to count as a causally efficacious property which plays a causal role for the entity instantiating it, that property or relation has to be identical to some physical property or relation.

---

[128]But Davidson is not advocating a 'strict-law' criterion of explanatory relevance. He says: "[It] is an error to think no explanation has been given until a law has been produced" (1963/1980:16-17).

I should also note that aside from the epistemology of causation, Davidson is an extensionalist about causation. It is only the explanatory nexus which is intensional and holistic (as per Fodor and LePore's definition of Holism), not the interaction itself.

[129]One example is a suggestion by McLaughlin (1989) of a model of mental causation which accepts the necessity of strict laws to subsume the physical, but relies also on "hedged"mental *causal* laws to subsume mental events. In this way, McLaughlin attempts to rescue Davidson from the charge of epiphenomenalism that his model seems to imply. For although Davidson requires strict physical laws as a condition for mental causal explanation, this does not mean, McLaughlin argues, that it precludes mental laws being causal. McLaughlin does not *endorse* this model however.

One of the major objections to this form of reductionism came specifically from within philosophy of mind in the form of the objection from multiple-realization. This objection, which fits nicely with the functional conception of the mental (see Putnam 1967; Fodor 1981), is based on the conceived possibility that higher-level properties and relations be *realized,* as functional types, by disparate lower-level properties and relations. A paradigm example that featured prominently in both Putnam and Fodor's writings from the period was of mental state types as computational types, contingently realized in some particular physical state or one reducible to it.

But notwithstanding the added finesse and interest that the computational model brought with it, this new form of materialism, non-reductive materialism, already found expression, albeit an implicit one, in Davidson's theory of mentalistic explanations. In chapter One we saw that Davidson conceived of mental events as basic particulars which in addition to their physical descriptions were also subsumed by rationalizing mentalistic descriptions non-definable in physical terms (Davidson 1963). This assumption, that the same token event could be given both a true physical description and a mental one, was the birth of the token-identity thesis and a foreshadowing of Davidson's solution to the problem of mental causation. Thus, in this paper Davidson in effect precipitated the ground for the claim that the truth-conditions of mentalistic causal explanations involve the identification of mental events, as basic spatio-temporal particulars, with physical events subsumed under strict causal laws.

In the philosophical literature, questions of inter-theoretic reduction are usually dealt with via the introduction of bridge-laws from kind terms of the reducing theory to kind terms of the reduced theory (E. Nagel, 1950; Shaffer 1963, Kim 1966). Since Davidson repudiated the availability of laws which quantify over mental properties (1970), and since in any case his identification of mental states and physical states is at the token and not the type level, he had to recruit a different notion to substantiate the latter kind of identification. He says:

> Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but different in some mental respect, or that an object cannot alter in some mental respect, without altering in some physical respect. (1970/1980:214)

This is how the concept of supervenience found its way to contemporary discussions in the philosophy of mind, and brought with it another wave of physicalistic sentiments. Thus, the idea embodied in the supervenience thesis of the mental on the physical gave rise to what is today called "the standard view"

about mental causation and mentalistic explanations. This is the view that beliefs, desires, and other mental states have to be brain states to figure in genuine causal explanations and predictions, even if as *state-types* they preserve their mental autonomy.[130]

The new brand of materialism differs from the old brand, the type-identity theory, in that it entails that the relation between mental and physical properties is not a necessary relation, or more accurately, that there is no necessitation from the instantiation of a mental property to the instantiation of some particular physical characteristics of the organism.[131] But it does require that the *realizer* of that state must be a physical state of the organism (although not necessarily an individualistic state).[132] To put it somewhat differently, while non-reductive materialism gave up on the requirement that mental properties be *identified* with physical properties, it still held onto the requirement, which to some might seem even more controversial, that mental states have "a non-intentional core", to use Hartry Field's picturesque phrase. I take it that something close to this idea is also expressed by Michael Tye in his demand that mental states "bear the same general ontic relationship to lower-level physical items as do the physical entities quantified over and referred to in higher-level physical laws generally" (1992:434).

The "neo-positivistic" picture that arises from non-reductive materialism on its various brands is of a structural model of mental states as having at their very being a physical "core" which is responsible for the driving force of mental states, or their causal powers. In this way, non-reductive materialism gave rise to the conception, as Kim put it, that "the mental does not enjoy its own independent causal powers" (1983:54). Thus, there is a sense in which the new conception of the mental, that of non-reductive materialism, has left intentional realists in a worse predicament with respect to the structure of the mental and its role in mental causation, compared to the type-identity theory. For on the new

---

[130]For a formulation of the standard view (to be criticized later) see Baker 1995, in particular ch. I. Criticism of that view can be also found in Wilson 1995.

[131]On Davidson's model of supervenience, weak supervenience as it was called and defined by Kim (1993, ch. 4), there is not even a necessitation relation from the physical to the mental, just a truth-functional implication relation.

[132]As Wilson notes (2001), the concept of realization, although introduced by Putnam more than 40 years ago (in Putnam 1960), still lacks systematic treatment. The same complaint was aired by Horgan (1993:573, fn. 20) who says that it should not be confused with supervenience. The supervenience base is much broader, Horgan says, than the realization base. I don't believe this difference matters much for my discussion.

model intentional realists are facing the risk of making the mental irrelevant as regards its own causal powers.

Perhaps somewhat paradoxically, it is this conception of the mental which I take externalists such as Burge (1979, 1982, 1986, 1989, 1992), Baker (1995), and others to take issue with. As we shall see in the next section, by advancing their anti-individualistic theses, Burgean externalists want to object to that metaphysical conception of the mind as occupied essentially by a non-intentional physical core, which putatively grounds the status of mentalistic explanations as a species of scientific causal explanations. My claim will be that the dispute over individualism that Burge initiated is thus first and foremost a dispute over the nature of mental states, expressed by the possibility of an alternative scheme of individuation independent of the neo-positivistic constraint.

From the above comments it seems that we are thus pretty much back at the issues that occupied us in chapter One, although from a different perspective. While in chapter One it was claimed that reasons cannot be causes because they belong to different logical"spaces", in this chapter the issue that forms the background will be that reasons cannot be causes since they belong in incompatible schemes of individuation. Reasons, as mental states, are individuated widely, whereas causes, as brain states, are individuated narrowly. My intention in the rest of this chapter will be to see in what sense mental states, as widely individuated states, can play a causal role in the production and control of behavior. Similarly, in what way physical states of the body, such as those arising from the stimulation of the senses (a prick in my arm), can lead to the production of wide thoughts (e.g., that John has just pricked my arm), thoughts which convey information about the environment. A centerpiece of the discussion will be Fodor's argument for narrow content as a sort of a scientific intermediate between the wide content of the Folk and the narrow physical states of individuals. The conclusion of this chapter will be that one does not need narrow content for mental causation since informational content is both truth-conditional, i.e., wide, and *compatible* with materialism.

2. Two Individualistic Theses

According to Burge, Individualism about mental kinds is the claim "that there is no necessary or deep individuative relation between the individual's being in states of those kinds and the nature of the individual's physical or social environments" (Burge, 1986:4, 1992:47). Anti-individualism is the position that there is such a deep individuative relation. Perhaps the most advertized thesis which makes such an individualistic commitment is the principle of Methodological Solipsism (MS), which was defined by Putnam (1975:220) as the taxonomic constraint that no ascription of a psychological state for the purpose of psychological theorizing "should presuppose the existence of any individual

other than the subject to whom the state is ascribed". In Fodor's hands this principle became the constraint that the psychological taxonomy of mental l states should make no reference to their strictly semantic properties, i.e., to their referential truth-conditions (1980:65; 1987:42).[133]

The principle of MS would be of course puzzling to those who think that psychological states are intentional states and hence, as per Brentano's definition of the mental (in chapter Four, section 1), *essentially* individuated by reference to their relation to objects and events in the world. But it is a sentiment that has adherents even in this post-Cartesian era. Thus Segal says that

> [I]n a taxonomy of thoughts that is appropriate for psychological explanation, the referent of a thought's singular component is irrelevant to the thought's type-identity. A viable taxonomy must collect appropriately similar thoughts about different things, and include in the same type even thoughts the singular component of which lacks reference altogether. (Segal 1989:40)

The reason for this conservatism of individualists in the face of what even they see as the forceful argument presented by the Twin-cases (on which below) is that, in their view, individualism is a somewhat different thesis than the one presented by Burge. That is, individualists such as Fodor (1980, 1987, 1991), Stich (1978, 1983), Devitt (1990), and others think that the issue that separates them from externalists is not in the first place that of the correct scientific taxonomies of mental states, but rather that about the metaphysics of causation. Thus, individualists for the most part see their position as making a substantive claim about the supervenience base of the causal powers of mental states, and only derivatively about their proper taxonomy. Thus, even as he concedes that the semantic properties of mental states cannot supervene on the intrinsic physical, Fodor claims that their *causal roles* must do so. He says:

> It is hard to see how internal representations could differ in causal role unless they differed in form. (1980:68)

---

[133]In his 1987 characterization of MS Fodor added the following in a footnote: "[MS] is a doctrine ...about individuation in aid of the psychology of mental processes. Methodological solipsism constraints the ways mental processes can specify their ranges and domains: they can't apply differently to mental states just in virtue of the truth or falsity of the propositions that the mental states express" (p. 42, fn. 9). Thus, it is a concern about mental *processes* that motivates Fodor's MS, by which Fodor means syntactic transformations.

And more recently he supplemented this with the following explication:

> Mental representations can differ in [their strictly semantic] content without differing in their intrinsic, formal, nonrelational, nonsemantic properties. But they can't differ in respect of the *mental processes* that subsume them except as they differ in their intrinsic, formal, nonrelational, nonsemantic properties". (In Fodor's reply to Perry and Israel in Lower and Rey 1991:298. Italics mine)

Thus, Fodor recognizes, in the spirit of the Twin-cases, that it is possible for the Folk individuation scheme of the attitudes not to be constrained by (what Burge calls) individualism, but objects that this individuative thesis cannot be inherited by the causal roles of these states, by which, as we shall immediately see, Fodor really means their causal powers. And indeed, it is *this* claim, *that the causal powers of a state of the organism must supervene on its intrinsic physical properties*, that I would consider to be the original thesis of individualism. In my view, it is this supervenience thesis which was then translated by individualists into a methodological constraint on the classification of mental states and their explanatory role. Thus Devitt has said that

> In psychology, we are concerned to explain why, given stimuli at her sense organs, a person evinced certain behavior. Only something that is entirely Supervenient on what is inside her skin -on her intrinsic physical states , particularly her brain - could play the required explanatory role between peripheral input and output. (1990:377)

The same idea was formulated by Stich under the *principle of autonomy*:

> The basic idea of the principle [of autonomy] is that the states and processes that ought to be of concern to the psychologist are those that supervene on the current, internal, physical state of the organism. (1983:164)[134]

it thus seems that we have *two* individualistic theses on our hands, one concerning the correct individuation scheme for mental states for the purpose of

[134]To be sure, neither Devitt nor Stich make here any mention specifically of 'causal powers' but only of mental supervenience. But below we shall see that a certain conception of taxonomic properties which I believe both Stich and Devitt accept makes for the connection between the two theses of mental supervenience and causal powers supervenience.

PEPs, whereas the other thesis concerns the metaphysics of causation. Fodor's own interpretation of these two theses came out as follows:

> Methodological point: Categorization in science is characteristically taxonomy by causal powers. Identity of causal powers is identity of causal consequences across nomologically possible worlds.
>
> Metaphysical point: Causal powers supervene on local micro-structure. (1987:44)

As we shall later see in more detail, the dispute between Fodor and Burge concerns the first point and in fact has almost nothing to do with the second point about the metaphysics of causation. The reason is that Burge does not think that psychological individuation of mental states has to be constrained by reference to the metaphysical base of mental causation but rather the other way around. That is, in Burge's view, questions of individuation constrained by working scientific theories are prior to the question which concern the materialistic grounding of causal powers. Thus, Burge thinks that Fodor errs by constraining the methodological issue by reference to the metaphysical issue. In a paper dedicated to the issue of mental causation he says:

> Our understanding of mental causation derives primarily from our understanding of mentalistic explanation, independently of our knowledge -or better, despite our ignorance-of the underlying [physical] process. Materialist accounts have allowed too wide a gap between their metaphysics of mental causation and what we actually know about the nature and existence of mentalistic causation, which derives almost entirely from mentalistic explanations and observations" (1993:103).

According to Burge, questions of mental causation should not arise within "a materialist metaphysical framework" according to which "only properties specified in the physical sciences are relevant to determining [causal powers] (ibid., 101), but should be dealt with rather within the intentional explanatory framework itself. Unlike Fodor and other "neo-positivists" who are concerned to establish the metaphysical basis for mentalistic causal powers, Burge thinks that the actual successful practice of mentalistic causal explanations should make the need for such a metaphysical basis superfluous. In fact, he thinks that if there is a problem with explaining the causal powers of mental states within the context of non-reductive materialism, this should be a sign that something is wrong with materialism rather than with the causal powers of the mental.

Given this set of convictions, one should take care, in my view, to distinguish Burge's kind of externalism from kindred positions which take the

causal issue to occupy a center role even within the wider framework of individuation. This in my view is the position of Putnam (1973, 1975), Davidson (1987:452), and Wilson (1995).[135] Thus, Wilson, in characterizing individualism has said that

> [individualism in] psychology involves abstracting over the mental states of *individuals*. It is concerned with identifying the cognitive contribution of the individual to her own behavior. As such, it is the *causal powers* of an individual's mental states that are relevant to psychological taxonomy and explanation. How an individual's states interact with one another, and how they, in turn, cause that individual's behavior are, after all, facts about that individual. Individuals form their particular mental states in various ways, but it is their being in those states rather than how they came to have them that is relevant to their subsequent behavior. (1995:7)

Here Wilson seems to recognize, I think, that the issue that individualists are concerned with is first and foremost that of the causal powers of mental states, where the taxonomic issue is derivative from it. In addition, it seems that Wilson takes the ways individuals come to have their mental states, mentioned as the anti-thesis to the individualist position, as empirical ways. For he takes the historical relations that agents bear to their social and physical environment as *causal*-historical relations. Below we shall see that Wilson takes such causal-historical relations of organisms to constitute certain *causal properties* which form the basis for the relevant causal explanatory practices. Because of such metaphysical concerns, Wilson's view of the contrast between individualism and anti-individualism is somewhat different from that of Burge. While Wilson thinks that the issue is between individuation by causal powers and individuation by causal properties, for Burge, as I just claimed, the issue is between two distinct *motivations* for individuation: the positivistic (i.e., materialistic) motivation, and the instrumentalistic motivation that he promotes. Because he accepts no constraints on individuation from the 'causal basement', so to speak, Burge sees no reason to think that wide individuation schemes do not individuate entities by their *causal powers* (this is quite clear in his 1993 paper). This is if the relevant explanatory scheme is a causal explanatory one, as FP clearly is.

---

[135]From my conversations with Wilson and from his writings I came to believe that he stands somewhat between Burge and Putnam/Davidson. On the one hand he seems to accept Burge's principle of the taxonomic priority of actual scientific individuation schemes, but on the other hand he still believes that one can keep the causal wagon rolling by substituting 'causal powers' with 'causal properties'. See below.

So to sum up the point that was made in this section, we saw that there were two individualistic theses concerning mental states. One, as traditionally understood by individualists like Fodor, Stich, and others, is the thesis that mental states should be individuated the same way that their causal powers are substantiated, which is by reference to their subvenience base: the internal physical properties of the organism instantiating those mental properties. The other individualistic thesis, as formulated by Burge, concerns the metaphysics of individuation as an independent issue from the metaphysical issue underlying causation or causal powers. Indeed, it is consistent with this latter thesis that one advances an individualistic position which is both non-relationally and non-causally constrained, so long as this position follows from some actual working explanatory practice. This could be the position of a Cartesian dualist.[136] On the other hand, one can advance a non-individualistic position which is not Burgean in that the relational individuation just follows from certain *metaphysical causal constraints* on individuation combined with a view of a wide supervenience base for mental states. A typical understanding of such a wide supervenience base is one where mental properties supervene not just on the subject's physical constitution but on the physical constitution of her environment as well.[137] In the case of Fodor, I should note first his "official" definition of individualism (or what he calls 'methodological individualism') as 'individuation by causal powers' (1987:42). As can be seen, there is no reference here as to whether the supervenience base is narrow or wide. But then, since Fodor thinks that causal powers supervene on the physical, and he also thinks that it is an empirical fact that the physical turns out to be 'local' (i.e., narrow, intrinsic), individualism, even by Burge's own definition, follows. From this we can see that Fodor's individualism *is not* the thesis that either causal powers, or mental properties, *necessarily* supervene on the *intrinsic* physical, but that they so supervene only nomologically. Let me explain.

Independently of the modal status of the supervenience relation itself, as a relation between properties, there is an independent issue of how the subvenening properties are instantiated, whether widely or narrowly (which

---

[136]I'm not claiming that Burge advances such a dualistic position, only that it is consistent with the logic of his argument. To show that Burge is a dualist, one will have to show that he allows that a) the mental is independent of any physical facts, and b) that the physical facts are independent of the mental facts. In his 1982 paper he comes close to (a) when he says that contents corresponding to natural kinds are determined by the community even when they are false (say, determining the content of water as XYZ).

[137]This is the position of Horgan 1993 and Tye 1992.

means: are the states they subsume individuated relationally or non-relationally?[138]). Depending on this latter issue, one can end-up with a supervenience thesis which is either narrowly or widely committed, but whose modal status can differ from that of the supervenience relation itself (say, as between weak and strong supervenience in Kim's terms (ibid.)). Thus, although the supervenience relation of mental properties on physical properties might be strong (i.e., necessary), if the physical properties are instantiated narrowly only contingently (in the sense that the physical laws might have allowed for their *wide* individuation), then the supervenience thesis of the mental on the intrinsic physical will itself be only with a *nomological* modal force, not of a *logical* modal force. But then the issue of the nature of the subvenience base, whether it is narrow or wide, will turn-out to be an empirical issue.

Now as we shall see in more detail below, according to Fodor, there is no individuation scheme of physical states of organisms which are individuated widely but still by reference to their causal powers . However, I will claim that there *is* such an individuation scheme of brain states which is one given by informational theory. In which case, one can advance a relational individuation of mental states which nevertheless satisfies the supervenience thesis of mental properties on physical properties, i.e., on widely individuated brain states. With this clarification behind us, I believe I can now turn to the Twin-cases.

3. Twin-Cases

We saw that individualism, as traditionally conceived by individualists (Fodor, Stich, etc.), is the position where the mental states of organisms supervene on those organisms' intrinsic physical properties because the causal powers of these mental states do. This individualistic thesis came under attack by using a series of thought experiments I shall call "Twin-cases". The strategy involved in the Twin-cases is simple: Show that two individuals can be identical in all of their "current, internal, physical states" but still distinct in their intentional states. Since the supervenience of the mental on the intrinsic physical is incompatible with such a possibility, then individualism about intentional states must be false. This is because we saw that supervenience is defined as the impossibility of a change in the supervening properties without some change in the subvenening

[138]I emphasize this wording in the text since I disagree with those who think that we can individuate *properties* relationally or non-relationally. In my view, properties are not entities that can be individuated either way, since they are *ways things (individuals) are individuated* (even when we take properties in the material mode). Thus, it is always the *instantiation* of a property, the individual state or event, that is individuated, be it narrowly or widely.

properties, whereas the Twin-cases purport to show that this *is* possible.

Thus, Putnam (1975) devised a thought experiment in which there is a planet just like Earth in all physical respects except that the micro-structure of the stuff which has all the manifest properties of our water and is even called "water" by the inhabitants of this Twin-Earth, is XYZ rather than H2O. Putnam then claims that when an inhabitant of Twin-Earth produces the vocable 'water is wet', his words would mean, in the sense of 'be about', something different than when an inhabitant of Earth produces the same vocable. When the Earthling, let's call him Oscar, says 'water is wet', he means that water is wet, in the sense that his words are about water that it is wet. But when the Twin-Earthling, Twin-Oscar, says 'water is wet', he means that XYZ (or twater) is wet; again, in the sense that his words are about twater's being wet. And this is so for both Twins, Putnam claims, even if they cannot distinguish H20 from XYZ.[139] The point is, Putnam argues, that Oscar does not have twater in his environment, and hence it makes no sense to say that he says anything about twater by 'water', and similarly with regard to Twin-Oscar and water.

The conclusion of Putnam from this thought experiment was that since the semantic properties of the vocable 'water' changed while the physical constitution of Oscar and Twin-Oscar remained the same, the semantic (i.e., representational) property of 'water' does not supervene on the intrinsic physical. And as this strategy can be generalized to other natural kind terms as well as others (proper names, indexicals, terms for artifacts), Putnam concluded that the meanings of all such terms 'are not (fully) in the head' (in Putnam 1975). That is, Putnam claimed that although there might be one component of meaning which could supervene on the intrinsic physical, this cannot be referential (i.e., intentional, truth-conditional) meaning.

In addition to this negative thesis, Putnam has supplemented his theory with a positive thesis about what referential meaning does depend on. Because the only factor which has changed between the Twins was the physical environment, it follows that the individuation conditions on referential meanings make essential reference to the physical environment in which the agent has acquired his or her language. That is, from a supervenience thesis of the semantic on the intrinsic physical, Putnam turned to a supervenience thesis of the semantic

---

[139]The emphasize on the epistemic angle is important once we compare this view with that of Russell, expressed in his acquaintance principle. This is the claim that one cannot be acquainted with an object (or a meaning) without recognizing that fact, and without being able to distinguish that object from others.

on the environmental-historical.[140]

Burge's thought experiments (in Burge 1979) differ from Putnam's in three respects. First, he applied the same strategy to the content of intentional states, not just to linguistic meaning.[141] Second, he claimed that the individuation conditions on these states go beyond the physical environment and include social and linguistic factors (but compare the last footnote). Third, he claimed that his thought experiments show that *all* contents should be individuated by reference to these externalistic conditions, including those that Putnam would consider as the narrow non-referential content 'in the head' (see in particular Burge 1982 where he maintains that wide individuation schemes apply to the *de dicto* contents of thoughts, not just top their *de re*, extensional content, as seems to be the case in Putnam's theory). As can be seen, Burge's challenge to the individualist is much more radical than that of Putnam.

The Burge cases begin in that we are invited to imagine an individual, let's call him Tyler, who believes that only written contracts are binding. Aside from that false belief, Tyler has many true beliefs about contracts (that they can be legally enforced in a court of law, that he just signed a contract for selling his house, etc.). It would seem then that we would take Tyler to have the concept of a contract, even given his false belief about contracts (that oral contracts are not legally binding). Next we are asked to imagine a physical duplicate of Tyler who produces the same vocables as Tyler. What is different in this counterfactual situation is not Tyler but his linguistic environment. For according to the

---

[140] Although it is true that this is not a result which is entailed by the structure of Putnam's thought experiments. That is, the argument does not show that a change in content was contingent on a change in the physical environment. Indeed, as I implied, Putnam also has a social component for the determination of reference, which comes out in his idea of the division of linguistic labor (in Putnam 1975). However, it is not clear that Putnam sees this social aspect more than serving as part of the causal chain between the speaker and the object/kind. Another social component appears in Putnam 1973 (207-8), where it is stressed that the causal chain begins with the ceremony of naming the object/kind, rather than with the object/kind(-instantiation) itself. But since this social aspect is not expressed in the structure of the thought experiment ('imagine a world physically like ours in which the naming ceremony had been different', etc.) , one is led to believe that here Putnam assumes that another causal chain to the object/kind(-instantiation) underlies the meaning of the naming act.

[141] Although Putnam has connected the implications of his semantic-linguistic considerations to the content of thoughts as that component involved in the understanding the meaning of linguistic items. (See Putnam 1975).

dictionary of this imagined counterfactual society (i.e., the précis of their linguistic conventions), a contract is defined as a *written* legally binding agreement. According to Burge, then, our Folk intuition would tell us that Twin-Tyler does *not* have beliefs about contracts. That can be shown, according to Burge, when we see that Tyler's belief that non-written contracts are not binding is false whereas had Twin-Tyler expressed himself in a similar way, he would be expressing a true belief. So when Twin-Tyler expresses himself to the effect that oral contracts are not binding, it is not a contract belief which presumably caused him to say so, but perhaps a 'shmontract' belief, where shmontracts are defined as agreements which are legally binding only when written.

We see that *pace* Putnam's thought experiment, which only attacked the idea that referential content is solipsistic, Burge's thought experiment is directed to whatever meaning component is left in the organism's mind, and possibly even to whatever *psychological* component is in their head.[142] For example, while Putnam leaves room for narrow content of water and twater to be given in terms of the description 'the transparent, liquid, potable stuff that is found in the oceans and lakes around here', no such thing is the case in Burge's theory. The reason is that in Burge's thought experiment, such a description would count as a content of *anything* only if that is what the linguistic community takes it to be. And since it might happen that in Twin-Tyler's community 'water' is *defined* as XYZ, then there would be nothing that Tyler and Twin-Tyler could share with respect to water thoughts. Although they might be in the same physical state, and although they might entertain the same 'stereotype' (as Putnam called it), there is *no* notion of content they both share. This raised a serious problem for those who later tried to rescue individualism by substituting narrow content for the Folk wide, externalistic, content. In addition, it shows, again, that for Burge the causal issue in the individuation of mental states is not paramount, since clearly the linguistic conventions about classification and categorization take over (see also fn. 136). Because of such considerations I label Burgean externalism as

[142]Thus in a footnote which in effect concerns his 1979 position, he says that although coherent "[it is] mistaken to hold that propositional-attitude attributions non-rigidly pick out physical events...there is no good reason to believe the very implausible thesis that mental events are not individuated ("essentially" or "basically") in terms of the relevant propositional-attitude attributions...I reject the view that the same mental events (types or *tokens*) [my italics] are picked out under different descriptions in the thought experiments (1986:fn. 7 p. 15). What Burge does allow to be common across Twins are the material *constituents* of mental states. But this is just to say that he allows that the Twins are physical Twins which, to make the thought experiments themselves coherent, he has to.

synchronous or contemporaneous externalism (I consider Baker 1995 to belong here as well) , as opposed to the diachronic or *causal* externalism that Putnam and (I believe) Wilson endorse. For the individuation of mental states does not depend according to him in any essential way on explanatory causal relations to aspects of the *physical environment*, unlike that of *physical* states (see Burge's objection to Davidson's externalism in Burge 1993:106).

I will return to treat this subject below, but here I want to raise the question what would the implications be of taking those external individuation conditions to be *causal* after all. For example, we saw that Davidson identified mental events (and events in general) by reference to their causal liaisons. Later on he extended these causal liaisons to the histories and futures of events (in 1987:452).[143] In line with this, a position of causal externalism would make Oscar and Twin-Oscar's mental events distinct if the relation that they bear respectively to water and twater are causal relations, rather than synchronous relations.

The importance of this can be seen when we consider the implications that this kind of externalism will have on either the supervenience or the token identity theses. For example, under Davidson's individuation scheme for events, a *non-causal* externalist position, as that of Burge, would give distinct individuation conditions for propositional attitudes than it would give for the individuation of the physical states of the organism. For while the mental states, as per the Twin-cases, are assumed to be individuated relationally as per their content, the physical events would be individuated narrowly since, *ex hypothesi*, they lack a content (cf. also Seager 1992:445). This then would imply both the falsity of the supervenience thesis and the falsity of the token-identity thesis. I have already explained why that would be the case for the supervenience thesis. As regards the token-identity thesis, this one would fail once it is claimed that the mental and physical tokens themselves would be putatively individuated according to incompatible schemes (one wide, the other narrow). The mental events are individuated by reference to their intentional (Folk) individuation scheme, as follows from the intuition underlying the Twin-cases, the physical events by reference to their local causal liaisons.

Now as said, all that would be true for the case of non-causal externalism. But given *causal-externalism* such as that of Davidson, the picture changes drastically. Unlike non-causal-externalism which discriminates between the individuation scheme of mental states and physical states, causal-externalism does not: both mental and physical events are individuated by their wide causal

---

[143]In Davidson 1987 (p. 452) he says about physical Twins embedded in different environmental contexts that "there is *something* different about them, even in the physical world; their causal histories are different". For a discussion of these two different criteria and their interrelation see Seager 1992.

liaisons. The upshot is that once we accept a wide individuation scheme for mental events by reference to the differential causal nexuses that different environments exemplify, we should accept a similar scheme for token physical states(!!) But this in turn will imply that both the supervenience thesis of the mental on the physical and the token identity thesis regarding mental and physical events can be preserved (I believe this gives another reason for why Burge should object to causal externalism given his rejection of the token-identity thesis). The question which now arises is whether the extension of this causal externalism to physical types makes any sense. Below we shall see that it does.

In this context, I should also note that token-identity cannot be falsified merely by showing that mental *types* and physical *types* have incompatible individuation schemes. A single token event, as a spatio-temporal particular, can instantiate both an intrinsic property and a relational property, which is a universal phenomenon (for example, my pen instantiates both its intrinsic mass and its relation to me as its legal owner). To falsify the token-identity thesis, one would have in addition or independently to show that the *tokens* instantiating the set of intrinsic physical properties (mass, charge, location, etc.) on the one hand, and the environmental relations on the other hand, have incompatible individuation schemes (cf. Burge 1979:109-113, and 1993:105). In which case, the change in context (as from an H20 world to an XYZ world) would deprive the physical token of a mental token to be identified with, as the mental token from the H2O world does not even exist in the XYZ world. Thus we see that the refutation of the token-identity thesis requires the more radical view where the environmental context determines not merely the type-individuation, but the token individuation as well. However, I doubt that this more extreme thesis is demonstrated by the Twin-cases, which leaves it open, in my view, that the same token state would instantiate both a local property (being a neural state) and a relational property (being environmentally related).[144]

---

[144]Burge (1993:105) tries to support his argument against the token-identity thesis by reference to the principle: "No occurrence of a thought could have a different intentional content and be the very same token-event or event particular". On the face of it, this seems to be just Davidson's principle of the individuation of mental events by their *content* (Davidson 1987:452). But Burge's support for this principle is different from that of Davidson. His point is that "the system of intentional content attribution is the [only] means of identifying mental states and events in psychological explanation and in our self-attribution" (ibid., p. 110). Hence, any other mentalistic attribution would have to imply a different event. Whether his claim is true or not I think is an open question. But even if it is, we should remember that for quite some time the logical behaviorists tried to convince us that the mental states are just behavioral states just because we have

But to return to causal externalism, I think that *Davidson's version* of causal externalism is an implausible one because it implies that any two events with different causal histories will count as distinct events. To the very least, it will preempt any attempt to make counterfactual causal claims about events. But one can adopt causal externalism not by reference to factual causal liaisons, as the causal-historical theory has it, but by reference to "causal consequences across nomologically possible worlds", just as Fodor defined causal powers (see page 127). I suggest that if one is to adopt externalism, it is this kind of *counterfactual causal-externalism* which makes the most sense, and as we shall see, one which can also be consistent with the supervenience thesis.

4. Narrow Content

Not all philosophers have accepted that the moral of the Twin-cases was that the Folk's individuation scheme of the attitudes is non-individualistic, that is, that it distinguishes Twins as manifesting different psychological or explanatory kinds. For example, Brian Loar has claimed that one has to distinguish the common-sense *psychological content* that the Folk use in their PEPs from oblique or *de dicto* content associated with the 'that' clauses of ascription sentences. The latter kind of content, Loar maintains, has to do with the social conditions of content determination, whereas the former, psychological content, is individualistic and has to do with people's discriminative abilities. Thus in the case of the agent who believes he has arthritis in his thigh, his social content, the one we would express by 'Tyler believes *that* he has arthritis in his thigh' is common to him and to the doctor who knows that arthritis is an ailment only of the joints, as Burge claims.[145] But, Loar maintains, Tyler and his physician differ in their psychological content since they associate different beliefs with it, and hence involve different discriminatory capacities.

From this Loar concludes that as regards the explanation and prediction of Twins' behavior, one can substitute for the content expressed by 'that'-clauses, that is, the social content that *ex hypothesi* distinguishes Twins, the psychological content which would do the same work with regard to explaining and predicting their behavior. This we can see when we notice that the wide contents themselves can be substituted for one another without a change in their

---

no independent way to identify mental states other than through behavior. Certainly this should give us pause.

[145]The arthritis case is similar to the contract case, just that in the counterfactual community 'arthritis's is defined as a disease of the joints and the thigh, whereas according to our definition it affects only the joints.

explanatory and predictive force. For example, in the case of Oscar and Twin-Oscar, one can substitute in 'Oscar didn't get into the water because he believes it was cold', every occurrence of 'water' with 'twater', and that without changing the explanatory power of the sentence. This is because the concept of content which the Folk use in explaining and predicting behavior, according to Loar, is one which is sensitive to the way mental states are related among themselves and to behavior and surface stimuli, rather than their relation to the social or physical environment.

Loar's psychological notion of content is one we have already met in my discussion of functionalism in chapter Two, that is, content as conceptual role. But *pace* Putnam's stereotypes, Loar does not claim that his narrow psychological content can be *expressed* by reference to the phenomenal or manifest properties that are common to water and twater, (as in 'Oscar believes that the transparent, liquid, tasteless stuff etc.,is cold'). For were he to claim so, he would be liable to the same sort of criticism that Burge launched against Putnam and that I reviewed (in section 2). Rather, Loar is aware that his narrow content will have to be given without the use of 'that'-clauses (as those have been appropriated, so to speak, by the social usage), and hence indirectly by other linguistic and contextual means.[146]

Notwithstanding two basic problems that I recognize in Loar's paper, it incorporates a basic insight that I think even the most zealous of externalists should take to heart. But I will begin with the problems. As said, Loar offers a notion of narrow content based on inferential or conceptual role on the lines discussed in chapter Two. In that discussion I concluded that a notion of content relativized to systems of beliefs and desires is one which is inherently holistic and hence one which is liable to the problem of meaning comparability. Thus in Burge's arthritis example, Tyler and his physician would not really disagree, as regards their psychological content, about the kind of ailment Tyler has in his thigh since their beliefs do not share that content . But I think Loar's theory suffers from a more serious problem, this time an internal problem to his functionalist theory. For from the same discussion in chapter Two, we saw that the role of 'that'-clauses and their interlocking semantic liaisons was an essential ingredient in constructing a functionalist theory from a functional-role theory. This is because a functional theory needs to be supplemented by the a priori semantic constraints that only 'that-clauses and their semantic implications can supply. Hence, a notion of content which bypasses the oblique constructions that

[146]Indeed, Loar claims that it is only because Twin cases are rare that a more precise and expedient tool for expressing narrow content has not developed. But is it really the case that the sort of *situations* exemplified by Twin-cases are rare? For a different opinion, see Perry 1997.

ascription sentences introduce, as is supposed to be the case with Loar's psychological content, cannot be one grounded in functional role. Of course, one can define directly a notion of content based on inferential or conceptual role, but then this is not a notion of content which can be connected to behavioral outputs and surface stimulations, as Loar assumes.

However, and as said, there is an insight in Loar's theory which I think is too obvious to ignore. The insight is that for a wide range of behavioral outcomes concerning molecular Twins , it indeed seems of no theoretical importance which of a range of competing candidates for the content of Twins' thoughts we use in our theories, so long as we use one of them. As we saw in the example above, whether we used the content *water* or *twater* to explain and predict the behavior of Oscar/Twin-Oscar, our respective successes or failures would have been the same. In fact, it doesn't even matter whether or not we knew which Twin it was whose behavior we attempt to explain. Or to put it somewhat more egocentrically, it seems that as regards Twin worlds (in which the agents are duplicate though not the environment), it makes little difference if any for the *user* of the PEP-theory to know who *he or she is*. Here one can imagine a fantasy akin to Twin visitation cases in which the PEP-user is switched in her sleep with her Twin and transported to Twin-earth. Let's call them Scie and Twin-Scie respectively. Now my claim is that nothing in the explanatory power of the theory that Scie and Twin-Scie use would change so long as the physical facts between the two worlds remain fixed. But now we are in the following situation. Either we take the theories that Scie and her Twin apply to Oscar and Twin-Oscar as the same theory, or we take them as different. If the same, then it would be hard to see how their subject matter can be different, that is, how the intentional states of Oscar and Twin-Oscar can be such, given a minimal amount of scientific realism that I believe many externalists would be willing to adopt. For what better evidence one can have for identifying subject matter than that the same theory is fully able to explain and predict, with equal level of success (and at times failure) both types of phenomena?

Of course an externalist might claim that the theories *are* different, since after all they are about a different subject matter. Scie's theory concerns water thoughts and water directed behavior, Twin-Scie's theory concerns twater thoughts etc. But is that reply based on actual scientific practice?[147] For I press the externalist to come up with a real example of two *distinct* scientific theories which are recognized by their users to be substitutable for each other without

[147]Both Burge and in particular Wilson (1995) complain against Fodor that his arguments are a prioristic and that they go against actual scientific practice (mainly, Fodor's methodological point). It is now time to return the ball to their court.

affecting anything in their theoretical import. And of course, there is no such example to be found since scientists recognize that it is the explanatory power and internal consistency of a theory which matter (as well as questions of simplicity, the aesthetics of construction, etc., which are also "internal" aspects of theories); much less whether the predicates it uses are relational or not.[148] I will return to this point below where I review arguments that purport to show that there are legitimate scientific taxonomies which individuate entities by their extrinsic or relational properties.

According to Loar, the Burge and Putnam thought-experiments have not shown that the content that the Folk use for their PEPs is wide content. But his was a minority opinion.[149] In the view of most individualists (Fodor, 1982, 1987 ch. 2, Block 1986, McGinn 1982, 1989, Stich 1978, 1983), the case was made by the Twin-cases that the notion of content that the Folk use *is* wide.[150] The question then becomes what implication should this result have for scientific psychology. Here two revisionistic positions have been offered. According to one reaction to the thought-experiments, the Twin-cases demonstrate the futility of a scientific psychology based on a Folkish individuation scheme for the attitudes. After all, it would not be the first time that the way the Folk classify entities needs a revision when we discover that it does not match or satisfy our scientific theoretical constraints. Since what we need in psychology is an individuation scheme which can subsume Twins as a psychological kind (perhaps given the Loarian sort of considerations about the substitution of theories) , and since an individuation scheme which is based in FP cannot satisfy that desideratum, we should do away with the intentional altogether and attend instead to the "purely psychological". That is the position of Stich:

---

[148]To be sure, the same argument similarly shows that individualists should not insist on non-relational theoretical taxonomies, at least as far as the explanatory power of theories goes. But I have already claimed that it is the metaphysical underpinning of causation which motivates the individualists' campaign for narrow individuation.

[149]But see also Searle 1983 ch. 8, and Egan 1991.

[150]We need remember that this result was more than that 'some content, or component of content, is wide', since this was something that the truth-conditional theory of content already established. After all, we already saw that Russell, and especially neo-Russellians such as Kaplan (1978), Evans (1982) and McDowell (1977, 1986, 1994) advanced a notion of content which was 'object-dependent'. The difference is that the Burge and Putnam thought experiments have putatively shown that non-singular terms, such as kind terms, artifact terms, and others are wide as well.

> we expect a psychological theory which aims at explaining behavior to invoke only the "purely psychological" properties which are shared by a subject and its replicas. (1978:574)

When we enquire what Stich means by the "purely psychological", we find that it is syntactical properties he has in mind, which means that he endorses the syntactical interpretation of Fodor's principle of methodological solipsism. Other interpretations of MS take 'formal' to consist in the neurological properties of the organism (see e.g., P.M. and P.S. Churchland 1983), but the idea is the same. Semantical properties are irrelevant to psychological theorizing since they are incompatible with individualism and individualism is a constraint in psychology. From which followed the contention that the mental-intentional should be eliminated altogether from psychological theorizing and that we just do science: syntactic psychology, neurology, or whatever.

A similar sentiment towards the futility of the Folk individuation scheme was expressed by Fodor, but with different conclusions:

> Any scientifically useful notion of psychological states ought to respect supervenience; mind/brain supervenience ...is, after all, the best idea that anyone has had so far about how mental causation is possible. The moral would appear to be that you can't make respectable science out of the attitudes as commonsensically individuated. (1987:30)

Here we can see that Fodor takes the violation of the supervenience thesis as destructive for the Folk notion of content, wide content, but his response is not to dump intentional psychology but to supply it with a scientific notion of content that *does* satisfy individualism. The idea is to adopt the Fregean insight that every contentful state has *two* kinds of meanings, only one of which is reference. So even if referential, truth-conditional, meaning is shown to be psychologically irrelevant, there is still the other component of meaning, sense, which could supply the goods, provided it can be shown to satisfy the supervenience thesis.

So far in this work I have discussed one conception of narrow content in terms of functional role. That was the conception of content coming out of CRT (the "short-arm" version). A second conception of functionalist narrow content, but where the functional properties are syntactic/computational properties, can be extracted from Fodor's thesis of MS and the computational model of the mind. Again Fodor, in a passage anticipating his response to Perry and Israel:

> So long as we are thinking of mental processes as purely computational, the bearing of environmental information upon such processes is exhausted by the formal character of [mental representations]... (1980:65).

What Fodor offers here is not a vision of a purely syntactic psychology, as Stich has it, but still a vision of intentional psychology with content implemented in syntactic functional states. A similar view was expressed by Pylyshyn:

> The formalist view requires that we take the syntactic properties of representations quite literally. It is literally true of a computer that it contains, in some functionally discernible form...what could be referred to as a code or an inscription of a symbolic expression, whose formal features mirror (in the sense of bearing a one-to-one correspondence with) semantic characteristics of some represented domain, and which causes the machine to behave in a certain way (1980:114).

From Pylyshyn's formulation we can infer that it is the isomorphic relation between the structure of the formal-syntactic processes and the semantic relations between their intentional objects that makes the formal properties into content bearing properties. In general outline, the idea here is quite the same as that expressed by CRT, but with the proviso that narrow meanings are constructs out of the narrow syntactical-role of brain states. Still, on either the syntactical or the more broadly inferential version of functional role semantics, it supplies a conception of content which does satisfy individualism. This is because narrow functional roles, on either the syntactical or the simply causal-neurological version, are second-order physical properties of brain states which make reference only to what is "inside the skin" of the organism.

However, as we saw, such a conception of content suffers from two major problems. First, it is inherently holistic and thus leads to eliminativism. Second, to satisfy the constraint from the Twin-cases it has to be non-representational, in which case the question arises what exactly is it good for? If content is to be causally efficacious in the production of behavior and other mental states, then we might assume, this is because it has at least something to do with what is going on in the world to which the behavior is directed. But that seems to be something that narrow content either totally lacks or, if it has it at all, then this would be by sheer accident.

Although in MS Fodor advanced a notion of content which was non-representational like the narrow content of functionalism, he later came to adopt a new conception of content which was representational relative to a context. In this way Fodor tried to reconcile the FP conception of intentional states with the demands of individualism that the causal powers of such states supervene on the physically intrinsic.

It might help us to better understand what is involved in such a project if we take a look at an example by Stalnaker concerning the concepts of weight and

mass.[151] Weight is a relational concept in that the quantity of weight can be determined only relative to a gravitational field. Thus it can be compared to the Folk intentional content which is determined, we now agree, relative to a context (environmental, social or linguistic). The mass of an object is something like its narrow physical state in that it stays the same across contexts. Oscar and Twin-Oscar would differ in weight had Earth differed in its gravitational field from Twin-Earth, but their mass would stay the same. But what is the mass of an object independently of its different weights in different gravitational contexts? One way to look at it is to say that mass is just a function from gravitational context of an object to its weight, thus, Mass=f(x)→W (where x quantifies over gravitational contexts).

Fodor's revised notion of narrow content follows the same line of thought that narrow content is a function from context and thoughts to truth-conditions. Fodor explains that idea as follows:

> It's presumably common ground that there's something about the relation between Twin-Earth and Twin-Me in virtue of which his 'water'-thoughts are about XYZ even though my water-thoughts are not. Call this ...condition C. Similarly, there must be something about the relation between me and Earth in virtue of which my water-thoughts are about H2O even though my Twin's 'water'-thoughts are not. Call this ...condition C'. Short of a miracle, it must be true that if an organism shares the neurophysical constitution of my Twin *and satisfies C* it follows that its thoughts and my Twin's thoughts share their truth conditions. (1987:48)

The idea here is to express narrow content in an indirect way, without any specification of what it *means*, aside from its functional characterization. In this way, Fodor attempted to give an extensional criterion for an identity in narrow content, as a purely referential truth-conditional conception, but in a way which accords with the Fregean doctrine that content determines reference. But let's see how successful that project is.

As has been already noted (Stalnaker, ibid.), Fodor's notion of narrow content as a function from context to content resembles to some extent Kaplan's notion of character. When Oscar says "I'm 30 years old" and when Twin-Oscar says "I'm 30 years old" it seems that there is some component of meaning both utterances share. This, according to Kaplan, is the character of the indexical 'I' which is something like a rule for its use. Then there are the truth-conditions of the utterances which are different for Oscar and Twin-Oscar. What makes "I'm

[151]Stalnaker 1989.

30 years old" true when uttered by Oscar is that he is 30 years old, whereas what makes it true in the mouth of Twin-Oscar is that *he* is 30 years old.[152] The point is that although the truth-conditions are different, it is the common character of the indexical, that it refers to the speaker of the utterance, that get us there. Only *after* the sentence is evaluated with respect to this rule and the context that its truth-conditions can be determined.

But here it seems that the comparison between Kaplan's character and Fodor's narrow content breaks. This is because in the latter case there is nothing we can discern in the physical structure of agents which can give us a clue as to how the mapping from context to content is to be drawn. That is, while in the case of Kaplan's character of the indexical 'I' we know that we need to look for the speaker of the utterance to determine the truth-conditions of the sentence, so far as Fodor's theory of narrow content stands, we have no idea what to look for. After all, on the face of it, there is no reason why the content of water thoughts on Earth won't be Beer, on Twin-Earth Glass, on Cousin-Earth Vodka, and so on. That is, the point is that although it seems to be presupposed by Fodor that something about the internal states of Twins plus their context determines what their thoughts are about, we are not told what that might be. As Stalnaker made the point (1989:296), Fodor's way of narrowing content can be similarly applied to narrowing down any relational property such as *being three miles a way from x*, where x can be a burning barn in one context, a chicken-coop in another context and so on. But surely, nobody should think that there is something about the intrinsic location (assuming absolute space) that, given a context, would contribute to the satisfaction of that function. The function, and by analogy, Fodor's narrow content, seem to be determined purely externally, which does not leave much semantic role for narrow content to play.

John Perry presented a theory which bears some basic relationship to Fodor's theory of contextualized narrow content, called the theory of the 'essential indexical' (Perry 1979). According to Perry, indexical thoughts such as "I'm making a mess" must have a non-propositional, non-substitutable, constituent which could unequivocally explain one's behavior after coming to recognize the truth of the sentence. This non-propositional constituent, the 'essential indexical', had to be a brain state of the organism so as to connect "directly" to the agent's relevant behavioral outcome. It follows from this that the non-semantic part of the mental state did the causal work in the production and

[152]It might claimed that if the physical identity of Earth and Twin-Earth extends to history, as it usually taken to be, then if Oscar is 30 years old then this fact could also serve as a truth-maker for the utterance "I'm 30 years old" in the mouth of Twin-Oscar. However I take this to be an artifact of the thought-experiment, not a substantive semantic thesis.

control of behavior, whereas the strictly semantic content supplied the truth-conditions of the act, given a context of instantiation.

We can see that both Fodor and Perry's theories resemble the picture of intentional states as having a 'non-intentional core' that I discussed above. Because of that, both theories face the same difficulties. Like Fodor's notion of narrow content, it is not clear how the semantic content of Perry's essential indexical is determined. Specifically in the case of the latter, it is not clear how such a brain state plays the role assigned to it by the theory such that it would determine its own truth-conditions given a context of its instantiation. As I will claim below, only some sort of an informational relation that connects such physical states to their truth-conditions can realize this mapping from contexts to contents. But still, some might argue here that Fodor's theory of narrow content as well as other individualistic positions arise from the wrong assumptions about the relation between individuation and causal powers. So let's tackle this issue first.

5. On Narrow Content and Causal Powers

I will first give Fodor's argument from causal powers to narrow content and then discuss the premisses of this argument. The argument can be summarized as follows (in Fodor 1987, chapter 2):

A) Science in general individuates entities by their causal powers.
B) The causal powers of an entity *must* supervene on its physical properties. In the case of organisms, these include some of their brain properties.
C) The scientific individuation of brain states is by their local (i.e., intrinsic) physical properties.
D) The individuation conditions of mental states under the Folk taxonomy are relational (this follows from the Twin-cases).
E) Relationally individuated states cannot supervene on non-relationally individuated states (as exemplified by the Twin-cases with respect to intentional mental properties).
F) Intentional mental states as individuated by the Folk (i.e., widely), fail to supervene on brain states (follows from C-E).
From the above Fodor concludes that
G) A relational individuation scheme of mental states is irrelevant to scientific taxonomy (specifically from A, B and F).
H) A non-relational individuation scheme for psychology is required, one which is based on a kind of content which satisfies mental-physical supervenience: narrow content.

The way G is entailed from A, B and F, is as follows. A says that science individuates entities by causal powers whereas B says that causal powers must supervene on the physical properties of the entity we want to individuate. It

follows from this that science should individuate entities by reference to their physical properties, although B is neutral between relational and non-relational individuation. F tells us that intentional states under the Folk scheme fail to satisfy the supervenience thesis, because they are individuated widely, whereas the physical states of organisms, their putative supervenience base, are individuated non-relationally (i.e., locally). Hence, such intentional individuation is not by reference to the physical properties of the organism, and hence not an individuation by causal powers, hence not a scientific individuation. Thus the claim by H that a scientific individuation by narrow content, which does satisfy the supervenience clause, is required. Of course, such an individuation scheme does not *entail* individuation by causal powers, it is only consistent with one. But this is all Fodor needs.

Now, I will later claim that this argument by Fodor for narrow content though valid is unsound, as it rests on a false premise C. But I will argue for this only after I finish reviewing more closely the argument on its premisses, and after I disprove some of the widely recognized objections to it.

Premise A, which Wilson calls 'global individualism' (Wilson 1995, chapter II), is the same as Fodor's methodological point I discussed above. This is the claim that scientific theories in general individuate the kinds they subsume by reference to their causal powers. The idea is, one might presume, that scientific theories are in the business of explaining the kinds of changes and processes that their subject matter exhibits, and it is a reasonable assumption that such changes and processes are driven by the causal powers of those states (cf. Fodor 1987:34).

Perhaps a kindred motivation for the presumption that science individuates states by their causal powers is the idea that we need to distinguish real taxonomic properties, those corresponding to real changes in the organism, from those describing mere 'Cambridge' changes.[153] Here it is thought that genuine properties, those whose instantiations are involved in genuine changes in the object, are those which are defined by reference to their causal powers, an idea expressed by Shoemaker as follows:

> What makes a property the property it is, what determines its identity, is its potential for contributing to the causal powers of the things that have it. This means...that if under all possible circumstances, properties X and Y make the same contribution to the causal powers of the things that have

[153]A mere 'Cambridge change' is one which does not necessarily involve a change in the object, for example, becoming an uncle (after Shoemaker 1979, 1980).

them, X and Y are the same property". (1980:212)[154]

It follows from Shoemaker's condition that two entities, like our molecular Twins, cannot have the same causal powers while instantiating distinct genuine properties. This also implies that if Twins *do* have the same causal powers and instantiate distinct mental properties, then the mental properties are not genuine properties and hence not taxonomic kinds (assuming that science, at least, individuates only by reference to genuine properties).

It was claimed by Wilson (1995, mainly chh.2 and 5) that premise A about global individualism is false since there is ample evidence that science at times taxonomizes not by causal powers but by causal *properties,* by which Wilson means the denotation of certain "relational or historical predicates" (ibid., p. 34).[155] Wilson thinks that it is possible for a scientific taxonomy and taxonomy more generally to classify groups of individuals by reference to relational and historical predicates such as 'being a university graduate', 'being a pensioner' and perhaps less intentionally, 'being a victim of the Hiroshima bombing'. According to Wilson, these are all examples of individuation by causal *properties* rather than by causal powers, since they concern facts about the individual's history (such that her parents are victims of the bomb), rather then her own causal powers. Wilson thinks that this shows that it is false that individuation is always by causal powers. Because of considerations such as these, Wilson also expresses doubt regarding Shoemaker's criterion for the identity of genuine properties in terms of their contribution to causal powers (ibid. pp. 119ff). This is since the above examples show, according to him, that relational properties contributing to causal *relevance* have the same claim to being considered genuine like those contributing to causal powers. I will return to Wilson's argument below but now let us move on to some of the other assumptions.

Assumptions B and C are implicit in Fodor's metaphysical point (p. 127). I have taken care to separate them since I think there is some confusion in the literature with regard to Fodor's insistence on the supervenience of causal powers

[154]Compare also to Davidson's identity conditions for events that was given above.

[155]Wilson takes 'causal powers' to be essentially "forward looking", i.e., as related to the power of the entity to act on or affect other entities. I doubt that Fodor would accept this characterization. As we saw, Fodor defined identity in causal powers as identity in consequences across nomologically possible worlds. But 'consequences' can apply to the way an entity was affected in as much as to its ability to affect.

on the intrinsic physical properties of the organism, mainly in the face of his allowing some relational properties (such as *being a planet*) to count as genuine scientific kinds. As I claimed, this confusion follows from misunderstanding Fodor's metaphysical point as the claim that causal powers are *essentially* 'local'. But as we can see from theses B and C, it is only the thesis of the supervenience of the mental on the physical (thesis B) which makes an essentialist claim. Thesis C is quite empirical in character, leaving open the possibility that science could come to individuate brain states by their relational properties, provided that is shown, as per premise A, that such a relational individuation is by causal powers. As I have claimed, a relational individuation of brain states is a possibility if it is found that the brain states of Twins have distinct causal powers in virtue of their environmental relations. Below we shall see why Fodor thinks that brain states individuated relationally do not have distinct causal powers and why he is wrong to think so.

This interpretation of Fodor's supervenience claim also better coheres, in my view, with the interpretation of supervenience in general as a relation between properties rather than between things. If Fodor's thesis were the claim that mental properties supervene essentially on the intrinsic states of the individual, it would make mental properties supervene on things (states of the organism), rather than on properties. Instead, the claim of B still leaves open the question how the brain properties are instantiated (in effect, how brain states are individuated). Fodor thinks that in the case of the physical states of the organism, they are individuated non-relationally, i.e., locally, hence his individualism. But in my view he takes this to be itself a nomic fact (i.e., not a metaphysical one), one which would have to be decided by science.

Thus it seems that it is assumption B which carries the weight of the argument. To repeat, assumption B expresses the supervenience thesis that the causal powers of an entity must supervene on that entity's physical properties. What justification can one give for B? Kim (1993b:208) has suggested the following principle about the relation between the causal powers of mental properties (M) and physical properties (P):

> The Principle of Causal Inheritance: If *M* is instantiated on a given occasion by being realized by *P*, then the causal powers of *this instance of M* are identical with (perhaps, a subset of) the causal powers of P.

Now if we substitute 'supervenient on' for 'identical with' in the principle of causal inheritance, we get something which in my view is close enough to B.[156]

---

[156]The difference is that B is less specific than the principle of causal inheritance (the latter narrows down the subvenience base to the realizing properties) and B takes the dependence relation to be between properties rather

I will not get into the charge of epiphenomenalism that Kim raises for those accepting PCI (including presumably my adaptation of it) since it will take us away from the discussion. So now, with this positivistic doctrine in mind, let's see how Burge confronts thesis B.

Unlike Wilson, who attempts to salvage the causal relevance of relational scientific taxonomies, Burge seems to think, as I have already said, that the causal issue is incidental to the metaphysical issue concerning individuation. This is clearly shown by the way he responds to the charge that a scientific individuation scheme which does not satisfy the supervenience thesis can give rise to the possibility of "action at a distance", or of environmental events which ""affect" a person's mental events and behavior without differentially "affecting" his or her body". Thus he says:

> Th[is] reasoning is confused. The confusion is abetted by careless use of the term 'affect', conflating causation with individuation. Variations in the environment that do not vary the impacts that causally "affect" the subject's body may "affect" the individuation of the information that the subject is receiving, of the intentional processes he or she is undergoing, or of the way the subject is acting. It does not follow that the environment causally affects the subject in any way that circumvents its having effects on the subject's body (1986:16)

My understanding of Burge's point which bears on his position of individualism is general is the following. Burge thinks that individualism is neither a thesis about causal powers, nor a thesis about causal properties and causal relevance. Rather, Burge takes individualism to be a thesis about individuation conditions. He thinks that once one sees that there are operating causal taxonomies of mental states that are wide, questions about causal powers and supervenience have to be settled accordingly. In other words, Burge thinks that instead of the metaphysical considerations about causation determining the way mental states should be individuated, the wide individuation scheme of mental states should be taken as a constraint on the solution to mental causation.

In light of this, when Burge supplies examples of historical individuation schemes, such as the individuation of lungs and hearts in biology, battles in history, and continents in geology (in Burge 1986:14ff.), unlike Wilson, he does not attempt to see them as examples of individuation by causal properties. Rather, by accepting that causation is probably a local affair, Burge wants to

---

than between instances of properties. I don't see that these differences matter much for my discussion.

show that the metaphysical issue underlying causation is not a constraint on scientific individuation, *pace* Fodor.[157] It is because of this he allows that while causal interactions remain a local affair, as we just saw, the laws subsuming those interactions quantify over widely individuated states. This, in my view, is a classical Davidsonian move, distinguishing the way events and their interactions are described - the intensionality of the happening - from the happening itself independently of its description. And again, this just follows from Burge's view that one should settle on a particular individuation scheme *before* one can decide on its ontological grounding. In this respect, Burge's argument is no less a prioristic than that of Fodor, it just reverses the order of explanation: rather than the physicalistic supervenience base explain how mental properties can have causal powers, as is the case with the positivistic line I reviewed above, the individuation project will decide the supervenience issue. If one can supply kosher explanatory schemes which are relational, as Burge attempts to do, then this would be reason enough to object to the individualistic constraint on scientific taxonomies, in psychology and beyond. Here is the passage where Burge summarizes this point:

> What we know about supervenience must be derived, partly, from what we know about individuation. What we know about individuation is derived from reflecting on explanations and descriptions of going cognitive practices. Individuative methods are bound up with the explanatory and descriptive needs of such practices. Thus justified judgments about what supervenes on what are *derivative* from reflection on the nature of explanation and description in psychological discourse and common attitude attribution. ..it seems to me therefore that, apart from further argument, the individualistic supervenience thesis provides no reason for requiring (pan-)individualism in psychology (ibid., 18).

From this we can see that by making his argument, Burge has in effect preempted Fodor's argument for narrow content. Although Burge objects to the claim in premise A that science individuates by causal powers, his deeper objection is to the implication that one can use that premise as a *reductio* against wide content, as Fodor did. Another way of driving this point home is by saying that even had Fodor argued that scientific individuation of mental states should be wide

---

[157]He says: "Let us accept the plausible principle that nothing causally affects breathing except as it causally affects local states of the lungs. It does not follow, and indeed is not true, that we individuate states of the lungs ...as supervenient on the chemically described objects and events which compose them" (ibid.).

because the supervenience base is wide, Burge would raise (or should raise) the same kind of objection. In that respect, it is Burge who is the real advocate of the principle of the autonomy of the mental.

6. On Relational Taxonomies and Causal Powers

Either because he misunderstood Burge's paper, or for some other reason, Fodor responded to the various examples of genuine relational individuation from the sciences as follows:

> Relational properties can count taxonomically whenever they affect causal powers. Thus 'being a planet' is a relational property par excellence, but it's one that individualism permits to operate in astronomical taxonomy. For whether you are a planet affects your trajectory and your trajectory determines what you bump into; so whether you're a planet affects your causal powers. (1987:43)

Thus it seems that Fodor has amended his definition of individualism from individuation by causal powers to individuation by causal powers *or* by what *affects* causal powers. In this way his response to Burge seems to be that scientific relational individuation, like the ones Burge and Wilson gave, satisfy individualism if they affect the causal powers of the entities instantiating them. Against this it was claimed by Stalnaker (1989) that this way of formulating individualism makes it a weaker thesis than the first formulation since it is consistent with wide individuation in psychology. For that which 'affects causal powers' can be just the property instantiated in the organism's environment. Wilson, who continues Stalnaker's line, argues that Fodor's response commits the fallacy of equivocation (1995: 49). His argument amounts to the claim that if science individuates by causal powers *or* by what affects causal powers ,then one could not infer the supervenience of those taxonomic properties on the intrinsic physical. This is because many of those entities that *affect* causal powers are external to the object (and, I presume, independent of it). For example, in the planet case, part of what affects its causal powers is the mass of the star around which it rotates, but surely that star's mass is not supervenient on the physical properties of the planet.

Indeed, because of this sort of putative equivocation between a notion of causal powers satisfying the local supervenience thesis, and that satisfying the actual individuative practices in the sciences, Wilson claimed that Fodor's argument is invalid. For on the one hand, Wilson argues, Fodor uses a notion of causal powers that satisfies the local supervenience thesis, presumably because it excludes relational properties, and on the other hand, Fodor uses a notion of causal powers which satisfies global individualism, one which *includes* causally

relevant properties which do not supervene on the local physical. But since there is no notion of causal powers that can satisfy *both* the local supervenience thesis and global individualism, Wilson concludes, the argument for narrow content fails to hold.

But I think that both Wilson and Stalnaker are wrong. The reason is that Fodor did not introduce the auxiliary premise, that science individuates also by what affects causal powers, as entailing that whatever science individuates by has to supervene on the local physical. Rather, Fodor's point is the much more trivial one expressed already in his identity conditions for causal powers according to which "identity of causal powers is identity of causal consequences across nomologically possible worlds" (see Fodor's methodological point on p. 127). Thus, presumably, affecting causal powers just *means* affecting causal consequences across nomologically possible worlds. There is no essential reference to an external entity (that is, to the denotation of the description 'what affects causal powers') but rather to a *situation* where the causal consequences across possible worlds have changed. Hence, Fodor's addition of "by what affects causal powers" is not a qualification or weakening of the first premise to his argument but just a restatement of it. A scientific taxonomy which is sensitive to differences due to what affects causal powers is just a taxonomy *by* causal powers.

Because I know the opposition here is fierce, I would try to fortify intuitions by using the following analogy. Suppose we plan to go on a trip to the Canadian Rockies and stipulate that this will only be on the condition that the weather is fair, defined as: no rain, no snow, and above-zero temperature. We might just as well have worded the condition to say, 'we shall go either on the condition that the weather is fair, *or* that nothing obtains that negatively *affects* the weather to below zero temperature, rain, or snow'. What did the second disjunct add to the original condition? Metaphysically speaking, nothing. This is because, even if we decided to stay home because we knew there were low rain-clouds coming our way, this would be in accord with our condition *only if* they negatively affected the weather. But then it is the original condition which is sufficient for the task, the second disjunct adds only a predictive or some other explanatory tool.

This analogy also helps to highlight, in my view, the role of *being a planet* in affecting the causal powers of planets. Surely, that property does not affect the causal powers of planets in the same way that low rain clouds affect the weather. This role is reserved for the star around which the planet revolves. Rather, *being a planet* affects the properties of planets (i.e., chunks of matter) in the same way that *being a cloud which gives rain,* or *being a cold front,* affect the weather: it affects the *satisfaction of the condition* on being a *fair weather.* Another way to put this is to say that instantiating the property of *being a planet* is a metaphysically sufficient condition for instantiating the property of *having such*

and such causal powers, which is not *a causal* effect. In the terminology introduced by Cummins, these property relations should be subsumed by a property theory rather than by a transition theory (see the discussion of his *property theory* in chapters One and Six). Hence, 'what affects causal powers' need not make reference to an independent entity which would thereby falsify the supervenience thesis. For both the relational property (being a planet, being a cloud which gives rain) can be instantiated by the same entity which instantiates the subvenening base (the planet or the weather respectively).

To conclude then, I don't think that there is an ambiguity in Fodor's formulation of his argument since the notion of causal powers he uses for global individualism is just the same notion of causal powers used in the formulation of the supervenience thesis. Of course, one could claim, as Wilson does, that science does not just individuate by causal powers but by causally relevant properties as well, or as Burge does, that science individuates by explanatory properties more generally. But now it seems that the burden of proof is on Burge and Wilson to show that all these explanatory properties are not by "what affects causal powers", that is, *by* causal powers.

Indeed, let's look at Wilson's own Hiroshima example. Suppose that there is a sense in which it serves some explanatory purpose to classify individuals under the description 'being a victim of the Hiroshima bombing', and to treat this property as a projectible causal property. For example, we might assume that people who instantiate it would be more likely to develop cancer, and so watching over them in extended periods of time might serve some medical purpose. But surely, all that would be true just because the relational property 'being a victim of the Hiroshima bombing' affects the causal powers of these individuals. For isn't a development of cancer one way in which one's causal powers have been affected? Cancer can cause dying and dying surely affects the causal powers of individuals. On the other hand, compare this taxonomic property with some other relational property that might also pick the same group of individuals, for example, the property of having expressed the vocable "Nanio Tabe Masca" (which means 'when are we going to eat' according to my Japanese expert) the 1000th time.[158] The reason why shall be surprised to hear of any scientific taxonomy based on this description is that it seems to us to lack any significance with respect to the causal powers of this group of individuals.

---

[158]Cf. Stich's example of the industrial robot who has "successfully performed his millionths weld" (Stich, 1983:167). Stich also tries to make the point that the historical aspect of the description, that relating to the 999,999 welds that the robot has performed, is causally irrelevant. But of course, that would be true only if all these past welds did not have a bearing on the physical constitution of the Robot, as they surely did (Cf. Stalnaker 1989:309).

In fact, this point, that not any relational predicate picks up a scientific property, was exemplified by Fodor in the following parody:

> I have before me this gen-u-ine United States ten cent piece. ...What in this time of permanent inflation, will this dime buy for me? Nothing less than control over the state of every physical particle in the universe. I define 'is an *H*-particle at *t*' so that it's satisfied by a particle at t iff my dime is heads-up at *t*. Correspondingly, I define 'is a *T*-particle at *t*' so that it's satisfied by a particle at *t* iff my dime is tails-up *t*. By facing my dime heads-up, I now bring it about that every particle in the universe is an H-particle...thus! (1987:33)

The T and H properties in Fodor's example are relational properties, the relata being physical particles and Fodor's dime. Clearly, though such relational properties are well defined, they will not be found in the physics books. The reason for that is, as Fodor makes the point, that such properties do not make any contribution to the causal powers of physical particles, whereas a relational property such as *being a planet* does make such a contribution. What does it imply to say that this latter relational property affects causal powers whereas that of being an H/T particle does not? Fodor's answer is that for a relational property to affect causal powers, there must be some *mechanism* that underlies the relation, a mechanism which is apparently missing in the case of the H/T properties. Thus, as I see it, by causal powers, and by what affects causal powers, Fodor just means 'is underlaid by a mechanism', something that Fodor takes to be a condition on the instantiation of any genuine property, not just relational ones.[159]

For example, viruses are sometimes taxonomized in Virology by reference to certain structural properties such as the type and size of the nucleic acid, the size and shape of their protein coat and the lipoprotein envelope (see Fields 1990). Thus, the taxonomic property of *being a virus* is intrinsic to the organisms instantiating it. But one would assume that it would affect the causal powers of token viruses that instantiate it in virtue of a certain mechanism that makes for that instantiation. Presumably, such a mechanism would be specified at levels lower than that of the Viral - most probably at the biochemical and physiological levels - a specification which would explain what instantiating that property *consists* in. In the same way, Fodor thinks that instantiating the property of being a planet involves some such mechanism, one which probably makes reference to

[159]That is, since not every *intrinsic* property is causally relevant either, mechanisms need be assumed for the implementation of the causally efficacious intrinsic properties as well.

gravitational fields, kinetic energy etc. That as opposed to the case of the T and H particles where there is no mechanism we could think of which underlies the instantiation of the H and T properties. And now in the same way, Fodor thinks that there are *no* mechanisms which underlie the instantiation or taxonomy of widely individuated intentional states, such as water or twater thoughts. This is because he does not think that there are mechanisms that connect *brain states* to such features of the environment as figure in wide individuation schemes. And it is *these* mechanisms that would have to underlie the instantiation of *wide* mental states.

To wit, when we look at Fodor's argument for narrow content, we see that mental supervenience, the necessary condition for intentional states to have causal powers, can fail to obtain because 1) Intentional mental states are individuated relationally, and 2) Brain states are individuated non-relationally (or locally). And we remember that it is the failure of mental supervenience that is at the heart of the problem with the wide individuation of intentional states, not that causal powers essentially supervene on *local* physical states (if this notion is even coherent). Now, I already said that Fodor has accepted the moral of the Twin-cases that mental individuation is relational. Which means that the failure of supervenience depends now on showing that the individuation of brain states is local. For if it were the case that parallel to the relational individuation of intentional states, there is a corresponding relational individuation of brain states, the supervenience thesis expressed in premise B could still hold, and there would be no need for narrow content. So let's look at this possibility.

We saw that according to the Twin-cases, mental states are individuated relationally by reference to their respective environments. Thus, Oscar's "water"-thoughts are individuated by reference to water, and Twin-Oscar's "water"-thoughts are individuated by reference to twater. Now suppose that we do the same with their brain states. Each and every one of Oscar's brain states would get the prefix 'H2O', whereas each and every one of Twin-Oscar's brain states would get the prefix 'XYZ'. If such a scheme was scientifically legitimate, the supervenience of intentional mental states on brain states would be preserved, since for every mental difference in the Twins' intentional mental properties there would correspond a physical difference, whereas failure of supervenience occurs only when there is a mental difference *without* a physical difference. In which case, to repeat, no need for narrow content will arise since supervenience would be satisfied with the only notion of content operating is wide.

But Fodor objects to this maneuver since he thinks that it would be achieved by endowing brain states with causal powers they cannot have. To say that brain-states can be scientifically taxonomized by reference to their relation to H20 or XYZ in their environment is, according to Fodor, just like attributing differential causal powers to particles by reference to whether they are related to Fodor's coin in one of its positions.

Now Fodor does not try to justify his claim that there are no mechanisms that connect brain states to H20 or XYZ and hence I can see someone who might try the following reply. 'Why should we say that there are no mechanisms which connect brain states to the environment? After all, as an advocate of the informational theory of content (which I will discuss in the next chapter), Fodor surely recognizes that there *is* a mechanism which connects neural states to environmental features, this being the very mechanism which determines their *content* under that theory. Hence, a relational individuation of brain states is in the cards, which would make for the rescue of mental supervenience. Conclusion, there is no need for narrow content after all'.

However, I assume that Fodor's response to this would probably be the following. Even if we grant that there are informational mechanisms that connect brain states to the world outside the skin, such mechanisms would not subsume brain states as neurological *types* but as neurological *tokens*. That is, although we assume under the causal-informational theory that a class of brain states can stand in some causal-nomological relation to instantiated features of the environment (such as the instantiation of H20 or XYZ respectively), presumably this class cross-classifies with *neural* kinds, given the thesis of the multiple-realizability of mental properties. Otherwise put, to assume that the class of brain states which realize the informational-semantic relations to the environment maps onto a neural type is to assume type-physicalism, whereas the multiple realization theory would allow that token brain states of *distinct* neurological types can form the same informational relations to the environment. Conclusion, brain states, as neurological types, cannot be said to be widely individuated in a way which corresponds to the individuation of the corresponding wide mental states, which is the same as saying that there are no mechanisms which underlie their environmental relations under their neurological individuation. But it is the latter which is required to make the wide individuation of *brain states* scientifically taxonomic. That is, what is required, seems to be Fodor's position, is that we show that it is neurological state *types* which acquire causal powers through their relational individuation, not neurological states-tokens. Hence, since that cannot be shown, we need to conclude that the situation with respect to brain states is the same as that with the H and T particles: There are no mechanisms which underlie that relation of brain states to the environment, and hence no causal powers. Q.E.D.

Now, if brain states cannot have causal powers as relationally individuated, then as relationally individuated they cannot also ground the supervenience of the mental as relationally individuated. But as mental supervenience is *necessary* for the causal powers of mental states (PCI), then intentional states widely individuated could not have causal powers, and without causal powers there is no *intentional causation*. Hence, to rescue some form of mental causation, even if it is only a revisionist notion, Fodor suggests narrow

content as a notion of content that *does* satisfy supervenience. This is in fact Fodor's argument which I think is sound and valid, if one accepts the supervenience assumption expressed in premise B. So maybe it is time to turn to it.

7. On Causal Powers and Supervenience

Let's suppose that we accept Fodor's claim that brain states must be individuated narrowly for scientific taxonomy (i.e., in neurology) . Still, the possibility is left open that a relational causal taxonomy of mental kinds is possible just because mental supervenience is not a necessary condition for their causal powers. One way to argue for this is by showing that relationally individuated mental states can have distinct causal powers even as they are realized by Twins. To wit, the strategy is to show that there are causal differences between e.g., the state of being environmentally related to H20 and the state of being environmentally related to XYZ without differences in their physical realization base.

Thus, we can imagine a case where both Oscar and Twin-Oscar utter 'Bring me water!' in their respective contexts. Surely, what Oscar would normally get is water, whereas Twin-Oscar will normally get twater. And this shows, the opponent of the supervenience thesis claims, that the mental states that presumably have caused these behaviors, say, the states of desiring water and twater respectively, have distinct causal powers, although, again, there are no physical differences between the Twins.

But Fodor objects to this argument since, as we saw from his identity conditions for causal powers, he thinks that identity or distinctness of causal powers can only be evaluated *across* contexts, in a counterfactual manner. As McGinn put it, "For any particular causal transaction there must exist a power involved in that transaction that is abstractable and identifiable across contexts." (1991:578) The reason for that is that causation, as the Humeans see it, is not a singular affair but a general notion that abstracts away from many instances of its particular realizations. This implies that to evaluate the identity or differences in causal powers of the mental states involved in the example we need to position the Twins in conditions which abstract away from their context, and this we can do only if we put them in the *same* context.[160] But then we see that when we evaluate the causal powers of the Twins' states in the same context, under the same conditions, they are the same. For example, had Twin-Oscar been on Earth, uttering 'Bring me water', would result in his getting... water, just like Oscar, and not twater. From this it seems that the causal powers of our respective

[160]In chapter Six I will claim that this is not exactly the case. Certain changes in the context can be effected, those which can be "quantified-over".

"water"- thoughts are, after all, the same.

But the objectors do not give up. That only follows, they say, if the behavior (verbal in this case) is described non-intentionally. But describe it intentionally, as you should, and you get a difference again. For what twater-thoughts cause for Twin-Oscar is first and foremost *twater-asking* behavior, not *water-getting* behavior. Since we are referring here to the same event which on Twin-Earth should be described, *widely*, as a twater-asking behavior, positioning my Twin on Earth should not change that. Indeed, that we need to individuate behavior widely is consistent with the principle which we all seem to accept, that identity of effects implies identity of causes. But if my water-directed behavior counts as type identical with my Twin's twater-directed behavior, then we would expect that the same would hold of the respective mental causes. But "by assumption the Twins' attitudes differ a lot", concludes Fodor, in the name of his opponent.

Fodor's response is that if this argument shows that the mental states of Twins differ, then it would also show that their brain states differ too. That is because while Oscar is in a brain state that causes him to utter the form of words "bring water" which means *bring water*, Twin-Oscar is in a brain state that causes him to utter the same form of words which widely means *bring twater*. Hence if these are distinct behavioral effects, and if they are caused by the Twins' brain states as we assume (as materialists), then the brain states would have to count as distinct as well. But, Fodor protests, "I thought we agreed a while back that it would be
grotesque to suppose that brain states that live on Twin-Earth are ipso facto typologically distinct from brain states that live around here". (ibid., p. 37)[161]

But in fact, Fodor is making here a mistake of false advertising what was actually shown with respect to the causal powers of brain states. To wit, the argument with respect to the causal powers of brain states has not established that brain states do not acquire causal powers due to their environmental relations. Rather, what was shown was that environmental relations do not endow brain states with causal powers as *neurological kinds*, and hence that no

[161]Egan (1991) objects here that the inference from the difference in effects (behavior) to a difference in causes (brain states) is fallacious since it presupposes just what the externalist wants to deny: that there is no possible taxonomic and explanatory contextual characterization of events. For surely, Egan argues, the same physical phenomenon could count as one kind of effect under one context, and a different one under another context. Her example is of a voice recorder which sounds the same but would mean different things to different audiences. But as this example presupposes mental activity (in the interpretation of the sound), it is Egan which begs the question against Fodor.

such relations can have any bearing on *neural* interactions as such. But this still leaves it open that brain states acquire causal powers as belonging to some other kind, perhaps a mental kind(!). After all, planets don't instantiate the property of being a planet as chunks of mass, since that mass does not change whether they orbit a star or not. As chunks of matter, a planet and its physical Twins are of the same type. But still, they can be of distinct types in virtue of their differential orbits (say around stars of distinct sizes).

It is interesting that a similar confusion underlies Baker's *objection* to Fodor's argument (1995:ch. 2). Baker claims that if it is shown that the widely individuated states of Twins have the same causal powers, then this will be also true of planets. For replace a planet with a physically identical chunk of matter and it will behave just like the original chunk. From which she concludes that something has to give: Either Fodor allows that physical Twins have distinct causal powers, or he cannot allow that such relational properties as being a planet affect a difference in causal powers.

Now I agree with Baker that Fodor should allow that entities identical in their (intrinsic) physical properties have distinct causal powers, but her argument falters in the way it describes the planet scenario. For it is not the case that in the thought experiment we substitute two physically identical chunks of matter, but one planet for another. What makes the other "chunk" a planet is of course the existence of a 'mechanism', which in the case of the planet it is one it has in virtue of its very physical nature. And it is this mechanism which makes it behave like a planet, given the right context. But then, to show that this is the case with respect to mental states, or brain states, that they also can be classified relationally, one needs to specify the mechanisms that make them be of an environmental kind. In my view, our best shot here would be to look for the right kind of informational mechanisms that make for that environmental relation.

We might get a better understanding of that when we review Fodor's second objection. According to Fodor, accepting the above argument for the causal distinctness of semantic states could imply similar consequences for the causal distinctness of the H and T properties. I assume the idea is that it is possible to claim that T and H properties do differ after all in their causal powers provided we describe their causal liaisons by using the same wide terminology. This is to say that T particles interact only with T particles, even if transported to H worlds, for what is an H particle with respect to the H world is still a T particle with respect to the T world. And causal relations, we now assume, are rigid: If a T particle causes another T particle in the actual world, it causes that particle *as a T particle* in every nomologically possible world in which it exists. From which follows that "visiting cases" cannot demonstrate causal identity for H and T particles. But surely, it seems "grotesque" to assume that the upright position of Fodor's coin can affect the causal powers of physical particles. This is a *reductio* argument which seems to imply the same with respect to the causal

powers of environmentally individuated mental states.

However, this *reductio* argument obscures a possible difference between H and T properties and mental properties that I have already alluded to. For I have assumed that unlike the case with the H and T particles, there might be informational mechanisms that connect brain states with environmental events in the same way that there are such mechanisms for planets and perhaps other entities that are individuated relationally. In fact, I would even go as far as to predict that many (if not all) of the relationally individuated properties assumed to play an explanatory role in the sciences probably affect the causal powers of the entity instantiating it through some such mechanism.[162] But as to the case of brain states, suppose that Oscar's 'bring me water' behavior is distinguished from that of Twin-Oscar *on Earth* because they would expresses distinct widely individuated behaviors. Then we could say that Oscar's behavior was caused by a brain state with the relational property connecting him to water through some mechanism, whereas Twin-Oscar's behavior was caused by a brain state related rather to twater through the relevant but different mechanism. And then the only thing we need to do is to *identify* the property of being reliably related to water with the property of being a water-thought (given some further qualifications to be discussed in chapter Six). And similarly, we will identify the property of being reliably related to twater with the property of being a twater-thought; and it seems I have explained how widely individuated intentional states can have distinct causal powers.

The idea then is this. Mental states, which are widely individuated states, are in fact identical to brain states when these bear certain kinds of informational relations to their environment (plus the added qualifications). This sort of connection is had by those brain states in virtue of mechanisms that are specified at the neurological level. By assumption, some mechanisms are such that they connect brain states to instantiations of water, as is the case with Earthlings. But there are also possible mechanisms that would connect the same brain states to instantiations of twater. Because of this, the first kind of brain states, which can cross-classify with neural states, carry the information *water* when they are instantiated, the second sort of (possible) brain states, carry the information *twater* and so on. Thus, the same type of neural states, as in physical Twins, can instantiate distinct mental states, with distinct causal powers, since a state which covaries with water has distinct causal powers than one which covaries with

---

[162]It was argued against this that it implies 'pan-semanticism'. But as we shall see in the next chapter, informational relations are not, as such, semantic relations. I think few will argue against the claim that nature is immersed in informational relations, when they realize that any covariance relation is an informational relation.

twater etc. At least in some possible worlds, water but not twater would cause instantiations of 'water', those in which water is distinguished from twater. But since twater is not a nomological possibility, then in most nomological possible worlds Twins will have the *same* mechanisms, and thus there is no need for narrow content (cf. Fodor 1994). On the other hand, there is no action at a distance etc., or any other violation of physical laws, since we now see that wide mental states are nothing but token brain states which form a special kind as environmentally related. We still keep to the token identity thesis (as was argued in section 3), although we now allow that physically identical Twins can have states with *distinct* causal powers. The price, as regards the supervenience thesis, is indeed to make the mental supervene on tokens of the brain which are typed *widely*, by reference to the environmental relations. Given a mechanism that establishes the covariance relation between brain tokens and environmental features, there cannot be a difference in the mental states that *these* brain states realize. This also means that mental processes can be brain processes even if they are not neural processes. For surely a mental process that goes from , say, water beliefs to water desires, although *realized* by brain states, goes through not in virtue of those states falling under a neural kind, but as they fall under an informational kind, and hence (given the further qualifications), as an intentional mental kind. In this way we can see how mental properties can be supervenient on physical properties, and how mental tokens can be identical to physical tokens, but also how mental causation is not strictly *physical* , i.e., neural causation. This is just the opposite of what Davidson, Kim, and other neo-positivists wanted us to believe.

To conclude, then, my argument has been that although Fodor's argument for narrow content is valid, it is based on a false premise C which says that the scientific individuation of brain states is by their intrinsic physical properties. I have claimed that a relational individuation of brain states as relational states is possible if these states are individuated as informational states. As I have implied, whether this is the case or not is an empirical issue which depends of the discovery of informational mechanisms that underlie such capacities. This conclusion, I take it, is also in the spirit of Burge's position since it takes the wide individuation scheme of the attitudes that he and Putnam have argued for as a constraint on individuation rather than the other way around. Still, there are some issues and complications that need to be worked out here. Most importantly, we need to develop a model of the nature of the informational mechanisms that presumably underlie mental states, specifically, in what way they differ from other informational mechanisms. This will be the subject matter of chapter Six.

Chapter Six

Informational Semantic Theory

1. Preview of the Informational Concept of Content

In the last chapter I discussed two interpretations of Individualism in the philosophy of mind. On the one hand there was Fodor's interpretation according to which mental states should be individuated *by their causal powers* (Fodor 1987:42). We saw that by the addition of two premises - that the causal powers of an entity must supervene on its physical properties, and that the physical properties of organisms are instantiated locally (premises B and C in Fodor's argument respectively) - Fodor's individualism entailed 'localism': The idea that physically identical organisms (molecular Twins) instantiate properties with the same causal powers. From this arose the implication that all schemes of individuation of mental states which taxonomize molecular Twins as *distinct* psychological kinds might miss important causal generalizations. Hence, the conclusion of Fodor and other kindred individualists that the scientific taxonomy of mental states should be constrained by the intrinsic physical properties of the individuals having those mental states.

Then, on the other hand, there was Burge's somewhat different interpretation of individualism. According to Burge, individualism is the position where "there is no necessary or deep individuative relation between the individual's being in states of those kinds and the nature of the individual's physical or social environments" (Burge, 1986:4) . Thus, Burge does not think that individualism is a thesis about causal powers or their metaphysical grounding, but first and foremost a thesis about how to taxonomize mental states. Accordingly, Burge advances his anti-individualistic position which says that there is "a deep individuative relation" that implicates the instantiation of mental states by an agent and her physical and social environment. But I said that Burge's anti-individualism does not counter Fodor's individualism as 'individuation by causal powers', since the latter leaves open the empirical possibility of a wide relational individuation of mental states but still by their causal powers. Rather, it is Wilson's position (1995) which is truly opposed to Fodor's individualism, according to which mental states should be individuated not by their causal *powers* but by their causal *properties,* whether these satisfy the supervenience thesis or not.

My own view on the matter was that satisfaction of the supervenience thesis was necessary to explain the instantiation of causally efficacious properties

by any object, including psychological agents, as any such instantiation must be realized by some sort of mechanism. If the supervenience of the mental on the physical is violated, then unless the existence of *mentalistic* mechanisms is presumed (an oxymoron, perhaps), the relation of mentalistic instantiations to their underlying physical realization would become un-explanatory. However, in adhering to the supervenience thesis and its underlying rational, I did not mean to imply that we should not accommodate, at least to some extent, the point Burge made that questions of metaphysical grounding should be *constrained* by active classificatory and explanatory schemes. Rather, I thought that we could do both; that is, accept the supervenience thesis of the mental on the physical *and* the constraints arising from the conceptual and explanatory schemes of the Folk and scientific psychology. My view was that we could accomplish such a feat by changing the subvenience base of mental properties from neural properties, as non-relationally individuated states, to a class of brain properties which constitute informational relations. The result of such a model of instantiation of mental properties would be, I implied, that we could now account for the way *physically identical systems* instantiate mental intentional states with distinct causal powers.

Perhaps a word about how I see the meaning of *physically identical systems*. In my view, there are two senses in which systems can be considered physically identical. One sense, which is accepted by Putnam and possibly Burge (but not by Davidson 1987, as we saw), is that systems are physically identical iff they instantiate the same intrinsic physical properties. This is the sense in which Oscar and Twin-Oscar are considered to be physical Twins, although they causally interact with physically distinct environments. But there is also a second sense of physical identity in which standing in a certain covariance relation to the physical environment can make for a physical difference. After all, one way to identify properties is by reference to the laws they enter into, and *ex hypothesi*, Oscar's 'water' states, and Twin-Oscar's 'twater' states enter into distinct informational laws: the first interacts with instances of water, the second interacts with instances of twater. So in this sense, the Twins are *not* physically identical (and we need not assume that just *being* in a different physical environment makes for this physical difference, as Davidson does (ibid., p. 452), since my claim is that only nomic relations to the environment count).

From the above we see that although Oscar and Twin-Oscar can be considered physically identical in the first sense of physical identity, they can still have *distinct causal powers,* because of their presumed physical distinctness in terms of their nomic interaction with their distinct physical environments. In addition, although I have established the subvenience base of widely individuated mental states by reference to physical properties, I have not done so by advancing a wide base of supervenience, as some have suggested (Horgan 1993, Tye 1992). This is because I'm not claiming that mental properties

supervene on physical properties instantiated by the organisms *and its environment*. This is shown by the fact that it is still possible that there are worlds which are physically identical - individuals *cum* environment - but in which Oscar and Twin-Oscar instantiate mental states with *distinct* causal powers, because these mental states have distinct causal consequences *across nomologically possible worlds*. This is made possible by the fact that informational relations are nomic, and hence counterfactual; just like dispositions, they show their mettle even when they are not discharged. Thus, brain states can be widely individuated states, with distinct causal powers, even where they are instantiated in identical physical and social environments. In fact, as we shall see, it is necessary that it should be the case that brain states which 'realize' mental states could be caused to instantiate even in the *absence* of their content property, in order for them to count as *realizing* intentional properties (I have in mind the situation where twater causes 'water' to instantiate on Twin-earth). It is called 'the robustness condition' which will be discussed at length in this chapter. But here it will suffice to say that from this perspective, the Twin-cases are not an embarrassment for the individualist position (that is, the one according to which genuine individuation of mental states is by their causal powers), or a pressure to become revisionist (to introduce 'narrow content'), but merely an expression of the nature of intentional mental states, that they are robust.

Now in my view this model of supervenience of mental properties on physical properties
could give the best explanation to the expediency of wide individuation schemes that externalists have flagged in the last twenty years or so. My claim is that those externalists were almost right in their insistence that relational individuations of mental states are kosher for explanatory practices in science and beyond, but that they fell short in explaining *how* that could be the case. That is, I have argued that externalists could not explain how relational individuation of entities could account for the causal roles of these entities. Like Fodor, I also think that only individuation *by causal powers* could explain these causal roles, and this is what the informational theory supplies: a wide individuation of entities by their causal roles.

In this chapter it is left for us to investigate into the possibility that widely individuated brain states, as informational states of organisms, can supply the subvenience base for mental properties. We shall see that the construction of such a subvenience base is no simple matter, mainly due to the fact that in addition to the regular constraints on informational relations that such states satisfy, they also have to abide by the constraints imposed on intentional states. This in the main means the possibility of error, which in the context of the causal informational theory is called the robustness problem already mentioned: how to make for the possibility that informational states of a system could be caused

to instantiate by worldly events which do not constitute part of those state's content.

As we shall see, one of the central obstacles in solving the robustness problem is the desideratum we imposed on ourselves, F, that systems identical in their intrinsic physical properties should be able to instantiate distinct informational mechanisms. To show where I'm heading in this discussion, I will discuss now a common and quite 'simple' example of a system which instantiates states with informational contents, the thermometer. We shall then see what implication this model might have for 'human detection devices'.

The thermometer is a device which is able to encode information about the fluctuating room temperature because of some dependence relation between its own states ($r_{1...}r_n$) and the those of the room temperature ($t_{1...}t_n$). One way to express this dependence relation is by saying that there is no change in the states of the thermometer without a corresponding change in the temperature of the room. This is to say that the thermometer's states are about the temperature of the room rather than, say, about how many people are in it, because its states *supervene* on those of the room's temperature and presumably not on the room's population (although surely the room temperature can change with respect to how many people are in it, but the former does not *depend* on the latter). However, this dependency of the states of the thermometer on the room temperature cannot be enough to show that each and every instantiated state of the thermometer *carries information* about the current temperature in the room. This is since *any* change in the room temperature that corresponds to a change in the states of the thermometer would satisfy the supervenience clause. For example, supervenience of the thermometer's states on the room temperature would be satisfied if a change from r_1= 30 to r_2= 32 would be accompanied by a change from 30 to 31 degrees in the room temperature, from 30 to 33 degrees, and so on. Hence, for the states of the thermometer to carry specific information about the room temperature, a dependence relation of *each and every* one of its states on those of the room temperature is required, such that the thermometer would be in a state r_1 only if the room is in t_1, r_2 only if the room is in state t_2 and so on, *ceteris paribus*. Rubbing the thermometer with your hand would falsify that condition.

Indeed, let me say here a word about the use of *ceteris paribus* clauses in the context of the informational theory. On the one hand, all special-science laws use *ceteris paribus* clauses and hence there should be nothing special if informational theory uses them as well. That is, one should not expect informational theory to be able to discharge in advance all its *ceteris paribus* clauses, and define when the conditions of operation are informationally ideal. But on the other hand, the *ceteris paribus* clauses play a special and significant role in my version of informational semantics which relies essentially on the notion of "informationally ideal" conditions of operation. By this I mean, as we

shall see in more detain below, *the possible limits of cognitive inquiry.*[163] This is a more complicated notion of operating conditions than is usually assumed for 'hedged' laws (e.g., Fodor 1974) since our very ability to conceptually express those conditions is at stake. My only hope is that in what follows all this will become somewhat clearer. Still, I think I can say that much here, that a functional delineation of information bearing states could do a lot to distinguish those of their activations which are informational from those which are not. Thus, the hand rubbing the thermometer should not count as contributing to its informational content since it is not part of the functional description of the device. Similarly, an activation of a brain state which is usually implicated in the detection of a certain piece of information (say, the presence of cows) should not count as informational if it was activated through unconnected neural processes (which is possible; after all, it is still a brain state). That is because in this case, presumably, it is not activated *as* an information bearing state: its *functional role* has been changed. In particular, we might presume that it does not interact with other information bearing states and with behavior in a way fitting the functional characterization of mental states.

I have mentioned the condition that each state of the thermometer should depend on some particular state of the room temperature. Surely this requires more than that the thermometer's states should supervene on the room temperature. It entails, in addition, that for each state of the thermometer there should correspond some value of the room's temperature which is *necessary* for its instantiation (again, *ceteris paribus*:).[164] Because of this, the fact of the thermometer's being in some one of its states will become *sufficient* for the room being in a certain temperature, that is, informationally sufficient. It could then serve as an *indicator* for the instantiation of the corresponding state of the room, that which is necessary for it. But this indicative sufficiency should not be confused with causal sufficiency. Indeed, the causal sufficiency goes in the other direction in that the event of the room being in a certain temperature is causally

---

[163]I want to distinguish my notion of ideal conditions of operations from Stampe's notion of 'fidelity conditions' (in Stampe 1977). Stampe's fidelity conditions are conditions of normal or optimal operation which is applicable to designed systems. This account will involve us in circularity since one has to know what information a device is designed to encode in order to determine when its fidelity conditions are satisfied. I will try to show below that my theory avoids this problem.

[164]In order not to have to write '*ceteris paribus*' after every modal notion, let me just state that all those modalities which qualify informational connection should be seen as qualified by the *ceteris paribus* clause.

sufficient for the thermometer instantiating one of its states, the very same state for which that room temperature is a necessary condition for its instantiation.

Overall we get the following picture: The states of the thermometer depend on the room temperature in that they can change only if the room temperature exhibits some change. This is the supervenience thesis of the physical states of the thermometer on the room temperature, which determines the general property that the thermometer's states express: The room temperature. In addition, each value of the room temperature is both necessary and causally sufficient for an instantiation of a state of the thermometer. Its necessity is required so that the states of the thermometer should carry *specific* information about room temperature. The causal sufficiency is required so that the thermometer's states should be *determined* by the particular states of the room temperature, and that its behavior could be *causally explained* by reference to the changes in its environment. But aside from these modal subtleties, overall we get a definition of indication not unlike that of Stalnaker's:

> [An] object *indicates* that *P* if and only if, for some *a* in the relevant set of alternative states of the object, first the object is in state *a*, and second, the proposition that the environment is in state *f(a)* entails that *P*'. (1984:13)

(where 'f(a)' stand for the function which assigns unique states of the environment to states of the detection device).

Now let's see what will happen in cases where there are certain changes in the room temperature without a corresponding change in the thermometer's states. For example, suppose that while the temperature in the room changed from t_1 to t_2, the thermometer stayed at r_1. This should not mean that r_1 would fail, in this case, to carry information about the room temperature, only that the information it would carry is disjunctive information, provided that the thermometer would be in r_1 *only if* the room temperature is in either t_1 or t_2. This is something I shall return to below in my discussion of the disjunction problem. A corresponding case arises when two states of the thermometer coincide with the same room temperature. Here, again, such states of the thermometer could be informational if their instantiation imply the instantiation of that room temperature. If both states depend on the same room temperature, as I'm assuming here, then the situation somewhat resembles the coextension cases I discussed in previous chapters. However, unlike the cases that form Frege's puzzle, here the two states of the thermometer carry the *same* information. To make these states of the thermometer informationally distinct, we will have to make them depend on distinct room temperatures, as I argued in chapter Three. I will return to the informational solution to Frege's puzzle in the next chapter.

At this point I want to complicate my story a bit, by letting the thermometer do some work that will depend on its capacity to register

information about the room temperature. We want to see how the *causal powers* of the states of the thermometer might depend on its detection capacity. I turn the thermometer into a thermostat that is connected to a heater system. Its function is to turn the heater *on* when the room temperature drops below a certain limit, and turn it *off* when the temperature rises above a ceratin limit. What we need is a theory that explains and predicts the behavior of the thermostat system.

To be sure, there are two ways to predict when the heater will be turned on or off. One way, perhaps the more reliable way, is to attend directly to the intrinsic states of the thermometer. We can mark them up by reference to certain 'internal' parameters, such as the level of mercury in the glass tube, and then mark those values below which the heater is on, and those above which it is off. But suppose that the thermometer, as is sometimes the case, is buried inside the heater system such that we have no access to its internal parameters. In which case we can capitalize on the dependency of these internal states on the 'external' states of the room temperature to make our predictions. If we know, or discover, that when the room is in temperature t_1 this is sufficient to get the thermostat into the state s_1 etc., then we no longer need access to the thermometer itself. We can describe the correlation between the behavior of the heater system and states of the room temperature directly, not via the way the latter affects the states of the thermometer. But we should notice that in this case, we no longer make reference to the thermometer as an informational device about the room temperature but rather we relate to the room temperature as giving information about the states of the thermometer. The roles have been reversed.

I intend this example to highlight the point that as far as the informational theory is concerned, information is what is locally available to the system in terms of the array of dependencies on environmental features we have described above, and it is this dependence which matters for the encoding and then use of the relevant information.[165] However, when we *describe* an informational system by reference to these environmental features, we bypass the indicative properties of the system. In which case we no longer refer to the states of the informational

[165]By 'encoded' information I do not necessarily mean (referential) *meaning*. We shall see below why not every encoding of information equals representation. Thus, a system can encode a piece of information in virtue of its covariance-relation with a state of affairs but not *mean* that state of affairs. This is Cummins' example of a skin rash carrying the information of high levels of ultra-violet radiation but not necessarily meaning it (in Cummins 1983:70). On the other hand, since any covariance is in some sense the 'encoding' of information, if the behavior of a system is dependent on its informational capacities, it is dependent on its 'encoding' capacity as well. But I hope we can see why this model does not necessarily lead to pan-semanticism.

system as *encoding* information but as *encoded* by information. And clearly, this distinction is significant.

It seems to me that the fallacy of overlooking the distinction between states which encode information and states which are encoded by information lies behind certain externalist tendencies to downplay the role of the internal structure of explananda, including psychological explananda, in deciding on scientific taxonomies. Indeed, considerations of internal structure might not matter to the external description of a state, say in matters of its disposition to response, if the reliance on external parameters is as expedient. In which case we say that the internal structure can be quantified-over in its capacity as the occupier of a role of information carrier. We can then *identify* that role by reference to those external parameters. However, and notwithstanding this, when we go to explain how the system is able to *exploit* the information that is carried by its states, the internal structure does matter. Here we come to the description of the relevant mechanisms that underlie the capacity of the role occupier to instantiate the relevant informational states, and which is responsible for its causal powers. I shall discuss more of this later when I return to Cummins' property theory, this time in the context of the informational theory.

So far I have abstracted away from an important difference between thermometers and natural systems. The difference I have in mind lies in the fact that the former can accomplish their task in part because they were designed to bear the required dependency relations to the room temperature rather than, say, the room pressure. What the design factor contributes is not only the internal structure of the device but also its proper conditions of operation. If the thermometer was designed to work approximately at sea level, then, under those conditions and those only, its states would be about the room temperature. But place it at the bottom of the ocean and it would probably show something different. This is where the design itself enters into the content specification of the states of the thermometer.

Let's suppose that a certain neural structure in living organisms operates as an informational system about the instantiation of properties in its environment. In particular, I shall assume that this is a detector of the presence of water. As in the case of the thermometer, to achieve that capacity, a certain dependency relation between the states of the detector and instantiations of water in its relevant environment would have to obtain. More specifically, it will have to be the case that a) *any* instantiation of water in its proximity would become sufficient for an instantiation a state of which I shall designate as 'water', and b) that nothing else would. In accordance with my discussion, I call (a) the detection conditional and (b) the information conditional. As regards (a), if not every "ripple" in water in the agent's environment would cause the device to enter into a 'water' state, then its detection capacities will be compromised. As to (b), if the 'water' state of the device was also caused at times to instantiate by

non-water, the information it would then carry will not be specifically about water but about that non-water stuff as well. That is, if the device systematically enters into the 'water' state in the absence of water, then it cannot encode information specifically *about* water, even if there are times when water does cause its instantiations.

Now, how do you get that device to satisfy these dependency clauses? I have mentioned teleology before, in terms of the conditions for which the device was designed to operate properly. In the case of natural systems, though, reference to such optimal conditions for which the operation of the system was "designed" to operate (by "mother nature", as Dennett likes to say, (e.g., in his 1991b)), is quite vague and even question-begging. 'When are the conditions for instantiating 'water' normal or optimal?' we might ask. 'Well, surely, when it is *water*rather than anything else which causes it' seems to be the only forthcoming answer.[166]

Another possibility to secure the right sort of relations between the detecting states and their environmental values is by focusing not on the properties upon whose possible instantiations the device depends, but instead on the relations *between* those possible causes themselves. This is the solution that I would favor and will develop in this chapter. Let me now say a few words about it in the way of an introduction.

In my example, a brain state of an organism, denoted as 'water', is presumed to carry information about water due to its nomic covariance with instances of water in the agent's environment.[167] Well, presumably if the device *depends* on instances of water to cause its instantiations, that is in part because of some properties that water has, properties that are
sufficient to cause the device to enter into the 'water' state in the presence of water. I assume that this is then the kind of dependency relation that would make 'water' carry the information *water*. But then, it stands to reason that the detection device could be activated by anything other than water which happens to instantiate the same properties as water does when the latter triggers its 'normal' activation.[168] For example, if water causes the device to enter into

---

[166]This is in a nutshell Fodor's response to Millikan (1984, 1986, 1989, 1991), in Fodor 1990b chapter 3 and in Loewer and Rey 1991:293ff.

[167]It does not have to be the proximal or even contemporaneous environment, so long as there is a nomic connection between the two events.

[168]This is basically the structure of the account given by Dretske at some point to false beliefs (in Dretske 1981:208-9). But unlike Dretske, I do not assume that this similarity is necessary for producing false believes. As Cummins argues (1983:74), being falsely told that s if F is quite unlike s being F, but it can still

'water' in virtue of its color, liquidity, transparency, etc., what I will call its water-lookingness (or WL), so might Vodka in an Evian bottle, the reflection of light on a hot surface, and so on.[169] So it seems to follows that the dependence of 'water' is not necessarily on water but more generally on anything which has WL to it. Of course, water could cause instantiations of 'water' not just in virtue of its WL but also via its other properties; for example, in virtue of its chemical structure H2O. In which case, it would be harder to "fool" the device, though not impossible: We can surely think of a stuff which is not water but which has the right chemical structure to attach to the same detectors (or receptors) that recognize H2O.[170]

Then we might imagine also the following case. Suppose that there is a stuff which is just like water but which differs from water not in its chemical composition but rather in its subatomic structure.[171] First let me show why I think that this stuff, let's call it 'shmoter', is not water.

Consider a possible world in which the intelligent population recognizes water only by its chemical composition. That is, we might assume that the perceptual system of these creatures is 'microceptual' and hence works directly at the chemical composition of things. Then, one day a philosopher ("micro-Putnam") introduces a Putnam style thought experiment to this community in which they are asked to imagine a world indistinguishable from their own in all physical respects only that the stuff which *micro-looks* like H20 is not H2O. It is claimed that the micro-physical composition of this other stuff, shmoter, is not protons but, say, shmotrons, although it can make up a molecule which is chemically indistinguishable from H2O. I don't see any reason why the

---

produce in me the false belief that s is F.

[169]I'm skipping here over those possible causes of 'water' which themselves reliably connect with water, such as the sight of the tap, a picture of water, etc. All these are cases where it is *water* which is the actual cause. See below.

[170]Curare, the south American arrow poison, does a similar thing when it attaches to the acetylcholine receptors on the post-synaptic membranes, as does atropine, the drug used to treat victims of nerve gas and the insecticides used in agriculture (Carlson 1999). All these are cases where foreign agents "fool" a certain functional device.

[171]A ready proposal might be 'heavy water' or D_2O (deuterium oxide) in which the hydrogen molecules are twice as heavy as those of hydrogen. I take the considerations invoked in the text to show that heavy water might be presumed not to be a kind of water.

philosophical community in this world should not be impressed with the argument and with its essentialist conclusion that H2O is the stuff which has the same physical micro-structure as the H2O where *they* live. Since I don't see any principled difference between the *dialectical structure* of this thought-experiment and that of Putnam (*or* Burge), I claim that any essentialist conclusion that follows for the water/twater thought experiments (or contract/shmontract) should follow for water/shmoter. Specifically, I claim that if we accept that water is not twater, we should accept that water is not shmoter either. But then, what *is* water, and what does this further possibility imply about my theory of content? Surely we cannot continue indefinitely like this since we would be left without *any* concepts.[172]

Well, to be sure, my intention is not to raise questions about water or other natural and non-natural kinds but to say something about the *concept* of water as opposed to the concepts of those other things which are detected like water. Roughly the moral I think we should draw from the plausibility of reiterating Twin-cases thought-experiments is that the conditions of possession of concepts have to make essential reference to some discriminative ability to stop the regress, although not necessarily an ability for the subject herself or even for her community. Possible discriminability, given certain constraints to be specified below, is rather what we are looking for, and that without the need to presuppose any special conceptual capacities on the part of the subject or her actual or possible society (such as those involving in judgments or the evaluation of evidential support).

Since discrimination between property instantiations is built into the informational theory, we can translate the conditions of the possession of concepts to those of the determination of conceptual content as follows. In the case of the concept WATER[173], my claim is that given the indefinite number of ways in which it could be caused to instantiate by non-water stuff (twater, shmoter, etc.), its content would be determined by reference to the limit at which *it is selectively instantiated in humans given all actual and possible resources that humans can come by* (and still remain humans). By a selective instantiation I

---

[172]I'm not trying to claim that the mere possibility of shmoter implies its plausibility, and ditto for other such contrived cases. To be sure, shmoter is nomologically impossible, but so is twater. Another question is whether this argument and its iterative structure is compelling. In my view, it is as compelling as are the Putnam and Burge arguments.

[173]We recall that concepts are mentioned in upper-case letters. Thus WATER should not be confused with the content-bearing state 'water' which is a physical state of the organism (a brain state in humans).

mean where a property is instantiated exclusively in discrimination from all others, given the relevant constraints concerning the above mentioned human resources, resources which can in turn go from the trivial to the momentous.

For example, 'water' thoughts can be at times caused to instantiate by vodka in an Evian bottle. But clearly it is easy for us to correct that mistake by either smelling or tasting it, which are capacities most of us have. It would be more difficult to discriminate water from twater since we need to do the chemistry, but surely that is not beyond our capacity either. Then there are such cases as that of shmoter which would probably stretch even our current capacities in physics, although they are still in what we call the realm of the possible. But eventually we shall have to come to those cases where the differences between water and some water look-alike, say '*imp*water' (from '*imp*ossible to detect'), might be such that there is no nomological possibility for humans to make them out. It is not only that we have reached the limit of our psychological potential, in the sense in which a better sense perception might make a difference here, but the limit of every intellectual potential that is possible for humans to achieve (what I shall call our *cognitive* potential). In which case, it will always be true to say that when we instantiate 'water', it would also carry the information about that other stuff which is nomologically indistinguishable from water although, as per the generalized Putnam strategy we applied, it is not itself water.

Thus, my claim is that the content of the state of 'water' in our detection device depends, in an essential way, on the capacities of the systems in which it operates, humans we presume, to distinguish water from non-water, in a matter to be specified below. Thus, I claim that the determination of the content of 'water', as a state of human detection systems depends on their capacity to distinguish water from all those things which only look like water, have the same chemical/atomic structure as water and so on, but which are still not water because of *some* difference possibly-discernible-to-humans they bear to water. This also implies that when that capacity, *actual or possible*, reaches its limit, where the differences become nomologically indiscernible to human intentional systems, the contents of the detection states merge. As we shall see below, this will make my informational theory a species of response-dependent theories, but with a much broader modal scope. I will now try to put the story I have delineated above in more detail and hopefully supply it with better argumentation.

2. The Purely Informational Semantic Theory

The informational theory of content is sometimes mentioned in the context of certain assumptions about the role and contribution of content to the ability of intentional systems to cope with their environment. It seems like a reasonable assumption that for organisms to survive and proliferate in a world which is not

of their own making, they should direct their behavior and adjust their wants and needs by reference to the information they receive from it. The having of states which carry information about the world, and the ability to process those states in a way which preserves that information, would seem to endow its possessors with certain obvious selective advantages.

Studies in informational theory in the philosophy of mind developed as a by-product of the epistemic evaluation of information content of perceptual states (see Dretske 1969, 1979; Stampe 1975, 1977; Some cite Grice 1957, or even Peirce 1931, as originators of the idea of informational content). Informational semantics is the attempt to extend this non-conceptual notion of content to all intentional states.[174] The *purely* informational semantic theory constrains itself to constructing a notion of content by reference to terms borrowed only from informational theory.[175] This as opposed to versions of informational theory, like that of Dretske (1981, 1986) and Fodor (1987), which mix the informational account with non-informational factors, like events in the history of the individual. Still, since Fodor's later theory (in Fodor 1987:chapter 4 , 1990a, 1990b chapter 4, 1994) is the closest attempt I know of (i.e., besides mine) to formulate a purely informational theory of mental content, I will extensively draw on it in my discussion of the informational theory.

The first thing that needs to be said with respect to Fodor's theory is that it was not Fodor's intent to identify meaning with information. That is because, for a number of reasons, meaning *cannot be* information. I have already mentioned the disjunctive character of information content, and I will return to it below. But there are other reasons as well. For example, it is clear that while informational relations are quite rampant (after all, they obtain between *any* covarying properties), meanings are scarce. Non-natural meaning has to be distinguished from information if Grice's distinction between natural and non-natural meaning is to be preserved (see Grice 1957). This point connects with the phenomenon that a signal carries information about anything which reliably causes it, while that cannot be the case with its meaning. Meanings are more discriminative and hence scarce than information.

---

[174]The notion of non-conceptual content is due to Evans (1982:151-70). The idea of a non-conceptual content is of a state whose instantiation by the subject does not necessitate the subject to possess any concepts that might be required to express that content. For example, one can be in a perceptual state of seeing a cube without one's seeing *that* it is a cube, or to believe *that* one is thereby seeing a cube. (See also Peacocke 1983, 1989, 1991, Davies 1991).

[175]In fact, the only version of purely informational theory that I know of is the one presented in this dissertation.

Another example of a difference between meaning and information, one which poses a further problem for information based theories of meaning, concerns the fact that a signal carries information not only about that which causes it but also about everything which is *conjunctive*, or coinstantiated, with that which causes it. As a result, information relations cannot distinguish between necessarily *coextensive* properties, such as between being a trilateral and being triangular, or between necessarily *coinstantiated* properties, as that of *being a rabbit* and *being undetached rabbit parts* (for the latter problem, see Quine, 1960:chapter 2). Given problems as those mentioned in the last two paragraphs, the success of making meaning out of information content depends to a large extent on the ability to 'fine-grain' information relations.

I have said that the concept of information content is basically that of a dependence relation. This dependency is expressed as the nomic covariance relation between two properties that realize the capacity of either of them to carry information about the other. I will put it more formally under the Informational Condition as follows:

IC: S-events carry information about P-events only if the generalization 'Ps iff Ss' is a *ceteris paribus* causal law.[176]

IC expresses a *qualitative* conception of information content that for many seems the most relevant to that of meaning (below I shall try to show that this conception is right). In contrast to the qualitative aspect, there is also the quantitative aspect of information content which has to do with the *amount* of information that a state registers. Thus, the quantitative measure of information born at the source and registered by the receiver is a function of the reduction in possibilities from what could happen to what did happen (Dretske, 1981:14). For example, a choice of one employee out of eight for a certain task would carry twice as much information than if there were only four employees to choose from, since there are twice as many choices of employees that *could* have been made (four choices in the second case, eight in the first).[177]

---

[176]My definition somewhat resembles that of Dretske according to whom "A signal *r* carries the information that *s* is F iff the conditional probability of *s*'s being F, given *r* (and *k*), is 1." (1981:65) The difference being only that I follow Fodor in making the conditional probability akin to that of *ceteris paribus* laws and hence less than 1.

[177]As Dretske notes, how information is measured depends also on how the reduction in the number of possibilities is achieved. If it is done by reference to binary decisions (say by tossing a coin), then choosing one employee out of eight will encode 3 bits of information.

As in the qualitative aspect of information content, here also the amount of information that a state registers depends on what might or could be the case, as opposed to what merely is, or was, the case. In the qualitative aspect, it is not what caused 'x' that makes for its content, but what could or would have caused it, given the right circumstances. And the same is true for the quantitative aspect, as shown by the fact that not any causal interaction between p (the source) and 'x' would result in the flow of information between them. Suppose that from a collection $p_{1\text{-}4}$ of events at the source, p_1 has actually caused 'x'. As this stands, we cannot know how much information was transmitted from p to 'x' since we do not know what was the reduction in the number of possibilities at the source with respect to all those other events which could have caused 'x'. For example, if each of $p_{2\text{-}4}$ could also have caused 'x', then the measure of the dependence of 'x' on p_1 would be reduced by a factor of 4, as compared to the actual situation where only p_1 could cause 'x'. As I will later claim, for informational semantics the reduction in the number of possibilities at the source to cause instantiations of a symbol cannot supply a quantitative notion of content since the number of those possibilities is indefinite.

Another important issue concerning the notion of information content is 'noise', or what is sometimes called 'channel conditions'. This issue is also relevant to the evaluation of the subjunctive conditionals that underlie the flow of information. Suppose that the occurrence of a name on a piece of paper carries the information who is the employee that was chosen from the eight employees. Still, that is not going to be all the information that the note carries. In addition, the same piece of paper carries the information consequent on the reduction in the number of possibilities regarding the signal itself. This is because the transmittance of information regarding the employee could have been made instead by the use of a different kind of paper, a different notation, a different vehicle of transmission, and so on. That the signal took the shape it did, that it was produced in the way it was, embodies a reduction in the number of all those other possibilities and hence a source of information which is not part of what the signal is there to convey. For another example, a message delivered on the television screen would also carry the information that a cathode beam was being used, or that the message was read by a woman. But presumably, all that is not part of the intended information content of that message.

We can put all features of the signal which have to do with its methods of production under the category of the *mechanism* for its implementation rather than as part of its content. This distinction between mechanisms and content matters to the construction of the counterfactuals used in evaluating the subjunctive conditionals underlying the flow of information. To separate the two aspects of the signal, mechanisms have to be "quantified over", as Fodor puts it,

in evaluating the dependency conditions between source and symbol.[178] He says: "The conditions for [information] constrain the functional relation between a symbol and its referent, but they quantify over the mechanisms that sustain these functional relations" (Fodor 1990b:56). And it is important to note that by mechanisms, Fodor also means theoretical commitments (that is, the theories used to affect covariance).

Quantifying over the mechanisms in evaluating information conditionals can either mean *fixing* the mechanisms in counterfactual constructions, or letting them *freely range* over a set of values closed under the relevant implemented laws. I think that some presumed problems of informational theory arise from confusing these two ways of evaluating informational contents. I will get to this later, but here we can note that quantifying over the mechanisms that underlie the flow of information has the advantage that the mechanisms themselves can be semantic, as is the case with inferential mechanisms, without risk of circularity. For the content presumed by such mechanisms does not enter into the constitution of the semantic content they realize. This also allows for linguistic practices to be absorbed into the channel conditions, thus reconciling the purely informational theory with Kripke and Putnam's intuition that naming practices and ostensive definitions play a role in semantic content, although not a constitutive role.[179]

3. The Disjunction Problem

I come now to discuss one of the major obstacles in deriving the notion of meaning from that of informational content. The realization that a signal carries information about all of its potential causes poses a special problem for

---

[178]Quantifying over the mechanisms underlying information relations means that there is no reference to a particular mechanism as the one which has to be instantiated, only that *some or other* mechanism will do (given certain constraints at the "upper-level"). In fact, this is consonant with the idea of the multiple-realizability of the mental. Sometimes, instead of taking about the need to quantify-over mechanisms, Fodor just says that they have to be *synchronous*. By which he means, I take it, that they have to be fixed to the here and now. On this fixation of mechanisms, see below.

[179] According to informational semantics, the semantic role that naming trees and socio-historical chains play can be in principle replaced by any other informational channel. For example, it can be replaced by a set of mirrors organized in the universe in a way which would allow us to literally *perceive* distant historical and contemporary events.

informational semantics. That is because this feature seems to be incompatible with what looks like an essential component of meaning, the possibility of error or misrepresentation (Dretske, 1986; Fodor, 1987:chapter 4, 1990b:chapter 4).

The problem with the possibility of misrepresentation arises as follows: If a state ('x') means that which causes it (X), then we seem to be locked between the horns of the following dilemma: 1) If *only* Xs cause 'x's, then 'x' means *X*, but then all of the 'x's are veridical and there is no misrepresentation. 2) If a Y (which is not an X) also reliably causes 'x's, then it would follow that 'x' means (*X v Y*) rather than *X*, and then again there is no misrepresentation. So in either case we have no misrepresentation and hence no meaning.[180]

For example, consider the case where we assume that 'horse' means *horse* because horses reliably cause 'horse' instantiations. But to get wild (erroneous) instances of 'horse's we need also presume that non-horses, say cows on a dark night, could also cause 'horse's. But then, 'horse' would mean (*horse or cow-on-a-dark-night*[181]), rather than horse. If so, then tokens of 'horse' caused by cow on a dark night are still veridical, and thus we still haven't shown how misrepresentation or error can arise.

From my discussion of information content in the previous sections we can see why the disjunction problem is special to meaning. A signal which is reliably caused by horses carries information about horses, whereas were it reliably caused by cows on a dark night as well, it would carry information on cows on a dark night as well (even when it is caused by a horse). Because a signal carries information about whatever it is which would reliably cause its instantiation, error is not a constraint on information content. But that cannot be the case with meaning for which error is an essential component. This lack of analogy between meaning and information seems to show that the disjunction problem is in a sense the problem of how to make for the *difference* between the quite prevalent concept of information (what Grice called natural-meaning) and the more restricted concept of meaning.

According to Fodor (1987:chapter 4, 1990b:chapter 4), what the

---

[180]The notion of a reliable cause that is used here should be distinguished from the epistemic notion of a reliable cause used by Goldman (e.g., in his 1988:54, see also his 1979). For Goldman a reliable cause of a belief is one which does not tend to produce error. For me a reliable cause is one which is entailed by a (*ceteris paribus*) causal law.

[181]The hyphens are used to prevent a conceptual deconstruction of the property. Having a concept which means (HORSE v COW-ON-A-DARK-NIGHT) does not imply having the concept COW.

disjunction problem in fact shows is that meaning (as opposed to information) is *robust*, where robustness is the phenomenon where a semantically valuable state *could* be caused by a variety of causes, not just those which pertain to its content. Of course, this is not a constraint only on informational semantic theory but on any other candidate for a naturalistic theory of meaning. I will call this requirement on meaning the Robustness Condition, and define it as follows:

RC: For all X, if 'x' means X, then there is a Y (which is not an X) such that Y→'x' is a *ceteris paribus* causal law.[182]

For example, for 'horse' to mean *horse* it is necessary that non-horses, such as a cow (on a dark night), could also reliably cause instantiations of 'horse's (*ceteris paribus*).[183] Below I will discuss some of the complications arising from this proposal but here we can note that solving the disjunction problem amounts, in effect, to giving an account of the wild occurrences of symbols: those cases where symbols are caused to instantiate by something other than their content property.

One attempted solution of the disjunction problem, that of Dretske (1981, 1986), was to pin the distinction between wild and veridical occurrences of symbols on another distinction, that between learning and post-learning situations. Roughly, the idea is that in the life of the individual there is a period where some signals are selectively conditioned, say by the help of a teacher, and these are the ones which will later form the veridical set. For example, if during the learning period the individual was conditioned to respond to horses but not to cows on a dark night, then any instantiation of 'horse' on an encounter with cows on a dark night *after* the learning period would be considered an error. Clearly, this theory puts the burden on what was encountered and conditioned for during the learning period, which makes it a hybrid of the informational and causal-historical theories.

There are some problems with this theory, though. The first is that it seems unlikely that a principled distinction could be drawn between the learning period and what comes after it (for a similar objection see Fodor 1987:103). In particular, is the learning period restricted to a certain time in the individual's

---

[182]That robustness is a necessary condition for meaning seems to be overlooked in Fodor's original discussions of it (see Fodor 1987:ch 4 and 1990b:ch 4., mainly the definition on p. 121). Rather, he presents it as a sufficient condition on meaning.

[183]Namely, the conditions have to roughly match those under which horses also cause 'horse's (in this case, on dark nights). I will later refer to situations of cows causing 'horse's under conditions in which horses do not.

early development, and if so, does it mean that she cannot learn new concepts after that period ended? Or does each concept have a learning period of its own? Then there is this problem. Suppose that during the learning period the individual was trained not to respond to cows on a dark night and identify them as horses. How is it, then, that after the learning period cows on a dark night would still cause instantiations of 'horse's, something the theory requires? If the agent has learned not to identify any horse-looking creature with horses, what has changed after the learning period? But perhaps the more significant problem is that Dretske's theory would fail to produce a notion of content which is fine-grained enough for our purposes.

Let us suppose that all the horses the individual has encountered during her learning period are also livestock from uncle Joe's farm. Then if she was conditioned to respond to horses, she was also conditioned to respond to *being livestock from uncle Joe's farm*, which means that the content of her 'horse' thoughts are indeterminate between *horse* and *livestock from uncle Joe's farm*. Then suppose that after the learning period our agent is exposed to a cow from uncle joe's farm on a dark night, which prompted her to instantiate 'horse'. In which case, she does not commit an error, which means that Dretske's solution is insufficient. Of course similar result would accrue to all coextensive (or better, coextensive-during-the-learning-period) concepts, such as 'a horse' vs. 'a horse or 2+2=4', and so on.[184]

4. Fodor's Asymmetric Dependency Solution

According to Fodor, the way to distinguish veridical from wild tokenings of a symbol is by making the laws that govern the instantiations of the former more basic than the laws which govern instantiations of the latter. In this way, Fodor hoped to solve the disjunction problem only using the resources of the informational theory, i.e., laws and their relations. The result was the so called Asymmetric Dependency (AD) condition:

AD: For all Y (not identical to X), if Ys qua Ys cause 'x's, then the causal law Y→'x' is asymmetrically dependent on the causal law X→'x'.

First let me say a few words about the idea of Ys causing Xs *qua Ys* which is expressed here. This idea is basic to the metaphysical view of the differential

[184]Dretske's solution has some ingredients in it which can also be found in the teleological/functional theory which succeeded it (Fodor, 1984; Millikan 1984, 1986, 1989; Papineau 1993). All these theories involve an actualist clause and thus suffer from similar problems to Dretske's theory.

causal powers of properties, which in the philosophy of mind was expressed by Horgan (1989) as Mental Quausation. Unlike the reductive (Davidsonian) view that only particulars have causal powers, the non-reductive position that Quausation expresses is that things have their causal powers differentially by reference to the properties they instantiate. The claim is that this is what makes the nomic subsumption of events intensional. Since two properties from different "levels of description" can endow an object with different causal powers, how its causal liaisons are *described* makes a substantial difference to what we take them to be. In informational semantics the supposition that events can enter into causal interactions *as* states of a kind gains further significance since it is assumed that not all of the properties instantiated by an event constitute the content property (see the discussion on p. 174).

Now, cashing-out the AD condition in the terminology of possible-worlds, the requirement becomes that there is no nomologically possible world in which Ys (as such) reliably cause 'x's but Xs do not reliably cause 'x's; whereas there are nomologically possible worlds in which X's reliably cause 'x's, but Ys don't.[185] In this way, X-caused 'x's are distinguished from non-X caused-'x' without our having to say anything about 'x's causal history, only about its disposition to respond.

Before I continue with this, it is important to note that although AD helps to explain the possibility of RC - of how something other than the content of a symbol would cause its instantiations without generating a disjunctive concept - it is clearly not sufficient for robustness, even given the satisfaction of its antecedent. For there are situations where X's causing 'x' is asymmetrically dependent on Y's causing 'x' but which are still not robust. One example is of the causal chain Y→X→'x' in which Y would not have caused 'x' unless X did and

---

[185] A nomologically possible world is one in which the natural laws which hold in the actual world hold. Paul Boghossian has claimed (1991:71), that this condition is too strong, for it goes against the possibility that there are worlds in which 'x' would mean Y. To correct this, Boghossian has introduced the distance metric factor:

'x' means X if there is a world W such that 1) in W X cause 'x' and non-X don't 2) W is nearer to the actual world than any world in which some non-X cause 'x' and X's don't.

Now it seems to me that Boghossian's worry can be avoided by the stipulation that worlds in which only non-X's cause 'x's are *not* nomologically possible. This, of course, would make 'x' something like a rigid designator, since it implies that 'x' means the same property in every (nomologically) possible world in which it is instantiated. I will elaborate on this below.

not the other way around, but in which it is the case that any Y-caused 'x' is also an X-caused 'x'. But then Y cannot be the wild cause of 'x' since it is dependent on X, whereas robustness requires an independent non-X cause of 'x. This will become important later.

On the face of it, the idea behind the asymmetric dependency condition seems quite simple. The instantiation of a token symbol is assumed to be primarily controlled by the content property. Presumably such control is mediated via the operation of a mechanism which underlies the detection of the property by the organism. In which case, and as I have suggested above, it is plausible that instances of other properties, under certain conditions, might trigger the same mechanism. After all, mechanisms are defined over a range of operating conditions and might be activated 'improperly', so to speak, when those conditions are exceeded. Thus, with respect to perception, it is possible that the mechanism which presumably mediates the detection of horses, and hence the control of the instantiation of the HORSE concept, would be tuned to those features of horses that cows on a dark night also instantiate. The upshot of AD is in the implication that cows on a dark night would not have caused 'horse's unless the mechanism for the detection of horses was already in place. This is to say that cows on a dark night-caused 'horses' are parasitic on the ability of the organism to respond to horses.[186]

5. Property Theory and Implementing Mechanisms

I have already noted that mechanisms are quantified-over in the evaluation of informational relations. For the metaphysics of content, the important point is not *how* the law that 'x's are caused by Xs is realized, but *that* it is more basic than the law about Y-caused 'x's. This is the reason why Fodor has insisted that informational semantics does not imply any strong form of verificationism (1990b:119). Indeed, verificationism, the position that a concept's content is the *means* of determining its correct application, would be implied only if the mechanisms responsible for property detection made a constitutive contribution for content determination. Still, it seems that the only way to cash out the AD condition is by reference to an ability of the organism, in whom the informational

[186]So far there is no asymmetry between horse-caused 'horses' and cows on a dark night-caused 'horses' since the dependence on the mechanism goes also the other way around. To achieve that asymmetry, Fodor stipulates that the worlds in which only horses cause 'horses' are nearer to the actual worlds than worlds in which both horses and cows on a dark night do (1990b chapter 4). But as the discussion below will show, it is difficult to cash out this stipulation in terms of the underlying mechanisms.

state is instantiated, to distinguish between the properties controlling the instantiations of the content bearing state.[187] A world in which horses cause 'horse's but no cows on a dark night do seems to be a world in which horses are systematically *distinguished* from cows on a dark night.[188] If there is a law that X's are indistinguishable from Y's, 'x' would have to mean (*X v Y*) rather than either meaning *X* or meaning *Y*. To get a better handle on what this implies for informational semantics, I suggest we take a closer look at what Fodor sees as the implementation base of intentional laws.

According to Fodor, since intentional laws are not basic laws, they are in need of implementing mechanisms: mechanisms which are themselves governed by more basic laws (presumably all the way down to physical mechanisms and basic physical laws). The model Fodor has suggested is an adaptation of Cummins' property theory I have already discussed in chapter One, although with some important modifications. It is represented roughly in the following diagram (figure 6.1):

$$\begin{array}{ccc} F & \rightarrow & G \\ \downarrow & & \uparrow \\ Mf & \rightarrow & Mg \end{array}$$

Figure 6.1: The implementation of intentional laws by lower-level mechanisms.[189]

In the diagram, F and G stand for higher-level properties, as I presume intentional properties to be (e.g., the concepts WATER and FIRE respectively), and Mf and Mg are their physical realizers respectively. Let me say a few words about how this model fares with respect to Cummins' property theory before I proceed. We recall that Cummins distinguished between two kind of theories

[187]Below we shall see that at least in the case of natural kind concepts, Fodor has introduced a modification which is not purely informational, and hence not verificationist even in the weaker sense discussed in the text.

[188]Or it might be a world in which there are no cows, or dark nights, or no cows on a dark night perceived by the organism, etc. But our interest should be given to those worlds in which there *are* cows on a dark night but the organism is systematically able to tell them apart from horses (mere accidental discriminations won't do).

[189]One change from a Cummins' property theory that already shows in this model is that it is not assumed that Mf is a functional analysis of F. This will become clearer later.

that are used successfully in the sciences to explain phenomena, transition theories and property theories (in Cummins 1983:chapter 1). Transition theories explain changes in a system by reference to subsumption under causal laws, which is a model akin to the deductive-nomological model I reviewed in chapter One. But not all features of a system can be explained as transitions within it. A case in point is that of dispositions. For a system may have a disposition to respond to, say, the presence of items in its close environment without it ever undergoing any change, as when the item is not present. In effect, since causal explanations invoke dispositional properties, it seems that transition theories *presuppose* another kind of explanatory model to explain their disposition base. This is where Cummins' property theory comes in.

According to Cummins, if we look at actual scientific practices we shall see that "many scientific theories are not designed to explain changes but are rather designed to explain properties" (ibid:14). The question that the theory attempts to answer is not why the system (S) acquired some property (P), but "what it is for S to instantiate P?". For example, a property theory would answer questions such as what is it for gas to instantiate a certain temperature, pressure etc. Here the answer given is in the form of a 'compositional analysis' of S into its components and their modes of organization or alternatively, an analysis of the property instantiated itself (functional analysis if P is a dispositional property, property analysis if P is not dispositional). Thus according to Cummins, ancient atomism is a classical instantiation theory since it aspires to explain the macro-properties of objects (shape, color, rigidity) by reference to their micro-constitution. As to psychological explanations, if we take mental properties to be dispositional properties, as in the disposition of a state to encode information about its surroundings, then a functional analysis is called for, but still by reference to the implementation base of the property. In short, we need a specification of the underlying mechanism.

Now, let's return to my diagram 6.1 and make the supposition that it analyses the instantiation of states with informational content in a device which detects the instantiation of properties in its proximity. That is, we want to explain the instantiation of G in Mg by reference to its disposition to respond to the instantiations of F in the world. Since I'm not trying to explain the instantiation of F, I can afford to suppose that F is in fact sufficient for the instantiation of Mf. For example, we might presume that if F is water, then an instantiation of water might be sufficient for an instantiation of water lookingness (Mf), and so on. I hope this will become clearer later. For another thing, I'm assuming that instances of F *cause* instances of G and hence that there are causal laws that describe this transition, although more effectively I'm trying to get beyond this to an explanation of the general *disposition* by the system to respond to the presence of F by instantiating G. That is, I'm trying to supply an explanation of the *higher-level* causal-informational law between F and G in terms of a

mechanism that is specified at the Mf and Mg level. Still further, there is no hindrance to specify that mechanism itself in terms of another causal process from Mf to Mg. This just shows something that I think Cummins also accepts, that the line between transition theories and property theories is not all that well-defined.

So to repeat, the task before us is to explain the instantiation of G, a mental intentional property I'm presuming, by a system S, such as the brain. I'm also assuming that G is a relational property of S according to the wide individuation scheme of mental intentional states I discussed at some length in the last chapter. This of course complicates the task since it now also requires to answer the question of how a *local* instantiation of a property can *explain* its relational nature. On the face of it, informational theory seems to be suitable for answering that question since, as I have said more than once, it makes content dependent on the instantiation of properties in the world. But as we shall now see, due to the counterfactual nature of informational relations and the requirement to evaluate them under identical local conditions (to screen-out 'noise'), informational content threatens to be inconsistent with my results about the Twin-cases. For, as far as informational states are concerned, identity in intrinsic physical structure of the organism seems to imply identity in implementing mechanisms, which in turn implies identity in information content of the states of the system. This surely makes the justification for my adaptation of Cummins' instantiation model more urgent than before, although it appears that Cummins' model is our only way to show how RC and AD can be satisfied together.

To understand this a bit better, let's try to apply Cummins' model of implementation to the case of the nomic covariance of 'horse' with tokens of *being a horse*. But this time, I shall add the conditions of robustness (RC) and asymmetric dependence (AD). That is, I shall require that in addition to the control of 'horse's by horses, there would be another property, say that of being a cow on a dark night (henceforth, cdn) which also causes 'horse' instantiations. Given this, I assign F, on the one hand, to the property of *being a horse*, and I shall stipulate a W (for wild) to the property of *being a cdn*. G would be the concept HORSE whose instantiation the model aims to explain, and Mf, Mw, and Mg are physical realizers for F, W and G respectively. In the case of Mg, I follow our convention and refer to it as the type 'Horse' which, as I now assume, is controlled (though asymmetrically) by both horse and cdn.[190] This leaves us to

[190]As per my notation, 'Horse' stands for the type of state whose instances covary with instances of horses (and now also with cdn's). But presumably, each token of 'Horse', (i.e., each 'horse'), is also a token of some brain type, if the implementation is in a human brain.

determine Mf and Mw.

The point to note about Mf and Mw is that if it is a law that cdn's cause tokenings of 'horse' (RC), and that they do so only if horses do (AD), then it is reasonable to assume that this is because of something that cdn's share with horses which is in addition to their physical implementation.[191] To a first approximation, let's call it 'horse lookingness' (or HL for short). This property, which is presumably shared not only by cdn's but also by Twin-horses, undetached horse parts, etc., would, in turn, cause instances of 'horse' if that is how horses do (perhaps through some transducing mechanism). Altogether, it would seem that we get the following model:

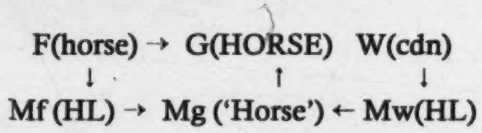


Figure 6.2: The implementation of the concept HORSE.

In figure 6.2 horizontal arrows represent causal laws, although at different levels of instantiation, and vertical arrows represent some variety of "bridge"-laws. I should note that although cdn nomically causes instances of 'Horse', it does not nomically cause instances of the *concept* HORSE. What makes it the case that 'Horse' instantiates HORSE and not (HORSE v CDN), (i.e., that there is no arrow between cdn and HORSE) is that the law cdn → 'Horse' depends on the law horse → 'Horse' but not vice versa.

Now, the causal laws marked by the horizontal arrows divide into two levels: the "higher-level" laws that subsume the informational flow from horse or cdn to HORSE, and those at the "lower-level" of description which subsume the implementing mechanisms for the informational laws. As I shall claim below,

---

[191]Some object that cdn → 'x' does not express a law in any case: there is no science that makes reference to such laws and it is not written in any of the acknowledged scientific books or journals. True, but the objector should note that the same was true for all currently acknowledged scientific laws at some point in time. Perhaps a better formulated objection would be that cdn → 'x' is not of the *type* to be recognized as a law, sine laws subsume kinds, and cdn's are not a kind. My response to this is that indeed cdn's are not a natural kind in biology or zoology, but so are left shoes, musical scores and end-of-the-year parties, all instances of properties with which I believe human intentional systems (and those alone), can *nomically* interact. Unfortunately, I will not be able to elaborate on this theory within the framework of the dissertation.

the range of the "higher-level" informational laws depends on the discriminative abilities of the kind of intentional system under consideration. As to the "bridge"-laws represented by the vertical arrows, those also divide into two kinds. First we have those "bridge"-laws that obtain between things (instantiation of properties and kinds) and some of their properties, such as their manifest or phenomenal properties. Cummins calls these 'nomic attributions' (ibid., p. 7), and we might presume that these cover laws between horses and their horse-lookingness, between water and water-lookingness, and even between cdn's and their horse-lookingness (I am not claiming that cdn's are nomologically horse-looking in the same way horses are, in fact, below we shall see that the AD condition *requires* that this is not the case).[192] Nomic attributions are distinguished by Cummins from 'instantiation laws' which presumably obtain between "lower" and "higher" level properties such as 'horse' and HORSE, 'water' and WATER and so on.

According to Fodor's AD condition, 'Horse' means *horse* and not (*horse v cdn*), if in any nomological possible world in which cdn's reliably cause 'horse's, horses also reliably cause 'horse's, but there are nomologically possible worlds in which horses do but cdn's don't cause 'horse's. But now, given my adaptation of Cummins' property theory, we can see that the cases where cdn's do not cause 'horse's can only be ones where (1) HL does not cause 'Horse' or (2) cdn is not sufficient for HL. Let's tackle (1) first.

I have assumed that cdn's cause 'horse's only in worlds where horses do and according to my model, that is because they are HL. So if horses cause 'horse's because they are HL, so should cdn's. This implies that if horses, or cdn's, cause (e.g.) Oscar, one of our molecular Twins, to instantiate 'horse' because they are HL, then they would do the same for any physical Twin of Oscar, such as Twin-Oscar. For the physical identity of the Twins suggests that they have the same detection mechanisms for HL. If I left it at that, we would have the result that informational semantics just amounts to the theory of narrow content where identity in physical constitution implies identity in informational content. That is because it now appears that for *both* Twins, the content of (e.g.,)

---

[192]Even if cdn's don't look as much like horses as horses do, *ex hypothesi*, it is still a law that cdn's look like horses if enough people would come to believe 'horse' when confronted with cdn's. If not, then I need to change the example. Some reliable covariance has to be presumed even in the case of wild causes since we don't want to conflate error with *exceptions* to regularities, which is an all pervading phenomenon. In this respect 'wild causes' exist in the shadowy area between exceptions and counterexamples. They cannot falsify theories as counterexamples do, but they cannot be discarded as meaningless aberrations either.

'horse' is the same, that is (*horse v cdn*), as both horses and cdn's would cause instantiations of 'horse'. But this implies that *pace* my promise in the first section that informational theory could supply a notion of content which endows Twins with mental states with distinct causal powers, it is actually inconsistent with a scientific taxonomy that individuates Twins as distinct psychological kinds.

We could derive similar consequences for the more radical cases exemplified by Putnam's water and twater. According to the Putnam's Twin-Earth thought experiments, it follows that water causes 'water' in Oscar iff twater causes 'twater' in Twin-Oscar (otherwise, behavioral and physical differences would be introduced). But if so, then by parity of reasoning this is because of something which water and twater share, that is, their water-lookingness (WL). It would follow, then, that in every possible world in which water causes 'water' in Oscar so would twater, since both are WL. In which case the content of 'water' for both Oscar and Twin-Oscar would be (*water v twater*), that is, the same content.

To avoid the result that a concept such as water must be disjunctive, a result which follows from the verificationist aspect of the purely informational theory (the need to be able to distinguish wild from veridical tokenings), Fodor tried to introduce an actualist qualification into the informational theory:

> Consider the Twin cases. Perhaps the first thing one is inclined to point to as relevant to distinguishing the WATER concept from WATER2 concept is that the former, but not the latter, is formed in an environment of H2O. But (purely) informational theories ...distinguish between concepts only if their tokenings are controlled by different laws. So if you want the WATER concept distinct from WATER2 concept, and you want to play by the rules of purely informational semantics, you have to assume a world where WATER is under the control of H2O but *not* under the control of XYZ... Correspondingly, the way you avoid the verificationism is... you let the actual histories of tokenings count too...(1990b:120)

The qualifying actualist clause that Fodor introduces to supplement the conditions for content is:

AC) Some 'x's are actually caused by Xs.

Given the addition of AC, it now follows from Fodor's theory that 'water' means *H20* and not (*H20 v XYZ*) as long as the actual history of uses of 'water' includes only samples of H20 rather than XYZ, as would be the case if there was only H20 on earth. A complication would arise, of course, if both H20 and XYZ were included in the actual history of uses of 'water'. In which case, Fodor thinks that

'water' would stop denoting a natural kind and denote rather a disjunctive property (in the same way that Jade was discovered to be disjunctive between nephrite and jadeite).

How reasonable is that suggestion? Well, my first reaction is that the problem is not unique to natural kind concepts as we saw in my discussion of horses vs. cdn's. If we cannot distinguish water from twater in any world physically identical to ours (or almost so), then we cannot also distinguish horses from cdn's, which would mean, given the actualist clause AC, that 'horse' means (*horse v cdn*). Second, we would need to face all over again the problems from the truth-functional aspect of the causal-historical solution. Suppose we want to determine whether 'water' means *H2O* or (*H2O v XYZ*), given that both H2O-caused 'water' and XYZ-caused 'water' are nomic, and that neither one depends on the other. According to AC, it all comes down to what actually figured in the causal histories of 'water'. Well, it so happens that H2O did, but since H2O is coextensive with (*H20 v XYZ*), then the latter also figures in the causal histories of 'water'. Conclusion: 'water' means (*H2O v XYZ*) after all.

According to clause (1) (on p. 187) , a world in which cdn's do not cause 'horse's is a world in which HL is not sufficient for 'horse's. Similarly, a world in which twater does not cause 'water' is a world in which WL is not sufficient for 'water'. But I have supposed that where horses cause 'horse's it is *because* they are HL, and where water causes 'water', it is *because* it is WL. So clause (1) cannot be satisfied together with AD since it suggests that a world in which cdn's do not cause 'horse's because HL does not cause 'horse's is also a world in which horses do not cause 'horse's. Similarly, a world in which twater does not cause 'water' is a world in which water doesn't. So there is no asymmetry between horses and cdn's or between water and twater. This leaves us clause (2) (again, on p. 187) according to which the failure of cdn's to cause 'horse's in some possible worlds is not because HL does not cause 'horse's but because cdn's are not sufficient in those worlds for HL. Similarly, according to clause (2) the reason why twater would not cause 'water' in some possible worlds in which water does is because it is not sufficient in those worlds for WL. Let's take a closer look at this option.

Suppose that in PW1, Oscar cannot distinguish horses from cdn's because both are sufficient for HL. But in PW2, Twin-Oscar uses some night-vision equipment and therefore can make it such that cdn's are not sufficient for HL. For looking at them through the night-vision equipment makes them look like cows, not horses. Or suppose that in PW3, Cousin-Oscar relies on experts to tell horses from cdn's (because *they* use a night-vision equipment, or because they can tell horse-DNA from cow-DNA, or whatever). I claim that all these are ways to make it such that while horses would still cause 'horse's in those worlds, cdn's won't, and this without a change in the *Twin's* physical constitution.

What I'm getting at is the following. We saw that according to clause (1)

we can affect the way cdn's cause 'horse's by making it such that HL does not affect 'horse's. That is since I assumed that cdn-caused 'horse's are dependent on HL-caused 'horse's. But then the result was that horses could no longer cause 'horse's as well since horse-caused 'horse's were also dependent on HL causing 'horse's. In which case, AD could not be satisfied since we have failed to produce worlds in which horses caused 'horse's but cdn's do not. But now I suggest that instead of affecting the way HL causes 'horse's, we should affect the relation between cdn's and HL (and therefore to horses). After all, HL, we might presume, is not an essential property of cows on a dark night, and so a property that they can fail to instantiate.[193] But then, affecting the relation of cdn's to HL should not bear on the physical relations between the Twins. To wit, the changes from PW1 in which both cdn's and horses cause 'horse's to PW2 in which cdn's don't cause 'horse's did not require any change in the physical constitution of Oscar or Twin-Oscar. That is since I'm assuming that the Twins' ability to recognize or be affected by HL as such has not changed. The idea is to effect a change in the world *external* to the Twins and that by *changing the ways that cdn's are related to horses* rather than the ways in which HL relates to Oscar or his Twin. Similarly, we can change the way water/twater affect the Twins not by changing the way WL affects them, but by changing the way water/twater are related to WL. That can be done either via the introduction of an instrument (perhaps an H2O/XYZ detector), an expert, or even the learning of a theory. Indeed, *even* by the use of theories since a theory is just a tool to affect discriminations between properties in the world which, like mechanical instruments, belong in the public domain.[194] Thus, the use of a theory can be one additional way where the relation between cdn's and HL, or between water and twater, could be affected without a change in the physical constitution of the Twins.

I think that we are beginning to see here the true significance of the claim of informational semantics that mechanisms should be quantified-over in assessing informational connections. For it allows us, as said, some latitude in the ways in which we describe the information channels that underlie these connections and in the way we evaluate the subjunctive conditionals that go to constitute information content. The point is that what has to be fixed, outside of

---

[193]In some sense it is also not an essential property of horses, but for that later. Here I would just ask the reader to accept that the relation of horses to their HL is stronger than that of cdn's to HL.

[194]Using a theory is pretty much like deferring to experts, who are, in turn, akin to 'human instruments'. This shows that a theory should not be presumed to be "in our heads" anymore than the experts, or the instruments, are.

the channel conditions, is only that which is *essential* to the informational link, and these are the source of information and the structure where it is registered. Everything else could be different than it is, so long as the flow of the *relevant* information (i.e., barring noise) goes uninterrupted. To repeat, what can be different includes the instruments we use, the experts we defer to, the theories we apply and even our own sensory make-up, as long as this is not where the information is encoded. A change in everything which belongs in the category of *mechanism* should not affect the relevant information of the signal, by which I now mean *its semantic content.*

Let's belabor our Twin-case. On both PW1 and PW2, Oscar and Twin-Oscar receive the information *horse* when they encounter horses (*ceteris paribus*). But on world2, Twin-Oscar fails to instantiate 'horse' when he encounters cdn's. The reason for that is that on Twin-Earth, Twin-Oscar uses a device, or a theory, or an expert that makes cdn's distinguishable from horses. Although this would affect the overall information that is registered by both Twins, that is, it would effect a change from (*horse v cdn*) to *horse*, it does not affect the information that *should* be registered: *horse*. This is another way of highlighting the difference between information content and *meaning*. The point is this. When we evaluate the counterfactuals that concern the information content at 'x', we need to fix all the physical facts about the Twins in order to get rid of the noise that accompanies information content. But when we evaluate the counterfactuals that concern only the *meaning* at 'x', we need to fix the physical constitution of the Twins, we need to fix the source of information, horses, but we can change the relation of cdn's to horse-lookingness. That is because the way cdn's affect us, or rather, the way we can distinguish between horses and cdn's, has nothing to do with the informational channel between *horses* and 'horse's. What it does affect is that when Oscar registers 'horse', he *means horse* and not (*horse v cdn*). For the sake of meaning, fixing *all* the physical facts goes too far. It is only the *relevant* physical facts which have to be fixed, the rest need only to be quantified over. Now, of course, there *is* a price to pay for this laxity and that is the introduction of more noise into the information channel. That is, instead of getting cdn's into the semantic content of 'horse', we get certain information about how to make cows look like cows rather than look like horses. But the upshot is that since there is an indefinite number of ways of discerning horses from cows, all those ways would eventually cancel-out (in the same way that all those indefinite ways of adding to the number 7 converge to the number). In the terminology of informational theory, we say that this sort of noise is background noise which does not 'step-over' the message.

6. The Theory of Informational Concepts

My purpose in the last section was to show that for informational semantic theory to work, the AD solution must not rely in an essential way on the detection mechanisms in making for the distinction between wild and veridical tokenings of symbols. As formulated by Fodor, the problem with the AD solution was that once we fixed the mechanisms in evaluating the subjunctive conditionals that constitute informational contents, we have thereby fixed the selective responses. My suggestion was to evaluate information conditionals by reference to worlds in which the relations between the respective external causes of mental tokenings themselves were affected. Thus, I have referred to worlds in which the relation between *being cdn* and *being a horse* has changed in that the relation of cdn's to horse-lookingness was affected, and thereby, indirectly, between cdn's and 'horse's. I claimed that in a similar way we can change the relation between water and twater via the effect on the relation of either to their water-lookingness.

Now, one consequence of this proposal is that the burden on the theory shifts to the metaphysical part which concerns the relation between properties rather than the relation between properties and minds. It also raises the question as to what our *concepts* of those properties would look like at the end of the day. In this section I would like to advance a metaphysical underpinning of informational semantics based on a kind of property which I call *a detection (or access) property*. I will argue that our informational concepts are determined by reference to our ability, actual and possible, to identify such properties via our selective responses to them. More specifically, I will claim that as far as my informational semantics goes, our *concepts* of natural kinds such as water and horse, as well as of non-natural kinds (pencils, symphonies, left shoes), are all metaphysical constructs out of such detection properties, given certain semantical constraints to be specified below.

In my discussion of informational semantics, we saw that there seemed to be an inconsistency between two demands that the theory puts on the notion of informational semantic content: the demand that the content of a state be determined by its reliable cause, and the demand that not everything which reliably causes this state should count as part of its content. To avoid the inconsistency, our task was to find a way to distinguish the wild tokenings of the content bearing state from its veridical tokenings. In the case of the state I referred to as 'horse', the assumption was that the veridical tokens are those caused by horses rather than by non-horses such as cdn's, because cdn's-caused 'horse's are dependent on horse-caused 'horse's but not vice versa. Then the question became, how to make it such that there would be possible worlds where horses cause 'horse's but cdn's, or other non-horses for that matter, don't.

My treatment of this issue was based on the supposition that the ways to make it such that nothing but horses would cause people, in a possible world, to instantiate 'horse's are equivalent to all those ways that they could *selectively detect* horses. These include, I suggested, both the actual as well as the possible means to affect such a selective detection. The actual means to selectively detect horses might include such common ways as that of telling horses by their particular look or by their characteristic whinnying sound; or alternatively, they could consist in the more sophisticated scientific ways of telling horses from non-horses by reference to their particular bone structure, their DNA composition, or any other means by which current science is able to accomplish this task when it is called to do so.

Outside of the actual means for the selective detection of horses, it makes sense to talk also about those means which are only a cognitive possibility from our current, actualist, perspective. What I *don't* have in mind here is a possible change in our psychological make-up that would enable us to distinguish horses from cdn's, perhaps a change that would eventuate in a better night-vision. This is because I take our current psychological make up to be a nomologically given, give or take a bit. Rather, by the cognitively possible I have in mind all those ways in which we could develop our scientific (and perhaps nonscientific) theories and instrumentation to affect selective detection of horses *given* our actual psychological structure. For surely, the more our technology advances and the better theories we develop, the more efficient we will become in distinguishing horses from non-horses, but without changing who we are (i.e., humans).[195]

Now, this talk about actual or possible means to selectively detect horses should not obscure the fact that eventually, our ability to do so is due to those properties that horses share. As said, these include their particular look, their bone structure, a certain DNA composition, and possibly many other properties that we have yet to discover. Above I called such properties of horses which help in their selective identification their *detection (or access) properties*. My claim is that in the last analysis, the informational-semantic content of the concept HORSE is just a function of our discriminative interactions with the horse detection properties to the exclusion of non-horse properties. And ditto for our other concepts of things.

The detection properties of things in the world and their interrelations are an ongoing enigma for the mind with respect to its concepts. From the perspective of informational semantics, this enigma is tantamount to the open

---

[195]The theories and instrumentation need not be particularly about horses but about anything which can then be applied to horses. This is the 'trickle-down' effect of scientific development.

formula relating to the number of possibilities at the source that could have caused a particular state to instantiate. We remember from my discussion above that the quantity of information registered by a structure is always a function of the *reduction* in the number of such possibilities at the source of information. The point is, though, that we are beginning to see that such a reduction, and hence the quantitative aspect of information content that goes with it, is more truly described as contingent on our nomological capacities for selective discrimination rather than on how these things are 'in themselves'.

We saw that our ability to register the information content *horse* in an encounter with horses was due, in part, to our ability to identify horses by reference to those of their properties which signal the presence of horses to our detection mechanisms. It was a main point in the discussion that not just any identification of a horse by such a detection property, or even a number of them, is conclusive, as it was acknowledged, even required, that non-horses could become indistinguishable from horses under certain conditions. I assumed that this was so if these non-horses possessed one or more of the detection properties which are 'originally' owned by horses. In the limiting case, we could imagine worlds in which there are creatures which possess all of the detection properties of horses but for one, say, relating to some non-phenotypic feature of their DNA. Such genotypic difference could be a difference in a detection property of horses vis-a'-vis these other creatures since, like horse-looking or horse whinnying, it would help in the selective identification of horses.

Now, it might be claimed that I haven't yet established that these 'Twin-horses' are not horses, and hence that I haven't shown that the said property of *horses*, in this case their particular DNA composition, is their detection property. Well, on the one hand, it would seem that such a difference *should* make a difference to the biological classification of these creatures, especially as these creatures might have a different evolutionary history from horses. But on the other hand, I don't think it matters for the *informational* concept of horses. What does matter for informational semantics is that there is a *possible* way for us to distinguish horses from these creatures, and reference to the kind of DNA molecules in their cells is such a possible way. Whether we should then decide to *classify* them as distinct natural kinds is another question, which carries, in my view, no semantic significance.

I will later return to this point but let me now move one step further and rehearse a case akin to the one I used in the first section. In a similar way to my previous example, we can imagine that there are creatures which are just like our horses but for the fact that the DNA molecules in their cells differ from those of the horses around here in some property that is *not possibly detected by humans*. Let us suppose that this is a property at the sub-sub-...-...atomic level which is nomologically inaccessible to us. What this means is that there is just no possible world accessible from the actual world where humans could detect that

substructure, and hence no possible world in which humans could distinguish these creatures from the horses here. To echo my question from section 1, are these creatures horses, and what does this possibility imply for our concept of a horse?

From the point of view of informational semantics, this case is just the other side of the coin of the case where we *could* detect the difference between horses and the Twin-horses. I have claimed already that if our concept of HORSE is informational, what matters is our ability, actual or possible, to distinguish horses from non-horses. We could make such a distinction in the case of Twin-Horses, and hence I concluded that they are not in the extension of HORSE. But this suggests that where we *cannot* make for such a distinction, as in the case of the sub-sub....atomic property, then our concept of this kind of creatures, '*Imp*horses' as I might call them, is the *same* as that of horses. Or at least, this is what my informational semantics implies. In metaphysical space, these differences might make for distinct kinds, but conceptually they do not. What makes this story palatable, in my opinion, is that informational semantics, to repeat, is not so much interested in the ontology of properties, that is, in how properties are in themselves, but in their nomic distinguishability for *humans*. Perhaps this is the source of the slogan that a difference has to make a difference, and I would add, even if this is only a *cognitively possible* difference.

Where I think this leaves us is that we now see that as far as informational semantics goes, verification does get into the picture about the individuation of concepts, but not in the way verificationists traditionally thought. The difference is that whereas for verificationism it is the *actual* methods or mechanisms for distinguishing between property instantiations that matter, in my theory the possible methods and means also count.[196] And this, to be sure, does extend the realm of possibilities *much* beyond what we usually think about the psychologically possible. For example, that we are able today to distinguish atomic and subatomic particles of different kinds of things implies the ability to manipulate some of the mathematical differences between them, but that capacity goes much beyond our physiological capacity to affect such discriminations. My claim is that it is such differences need to be included in the concepts of things even when we lack at present the appropriate theoretical knowledge to discern them. So it seems to me that even though my view implies a certain restriction in the way of carving out properties, that is, that our concepts will never be as fine-grained as the metaphysical possibilities out there, it

[196]Some verificationists, such as Ayer (1946:38ff) have talked about verification "in principle". However, as far as I can see, the idea of the *cognitively possible*, of the theories and instruments that we could possibly develop *given* our psychological make-up, has not occurred to them.

nevertheless sees our conceptual discriminations as going far beyond our psychological abilities (e.g., our perceptual abilities).

Let's turn back, for a moment, to the case of the concepts WATER and TWATER. 'Water', we presume, means *water* even though twater would also cause its instantiation in every possible world in which twater's chemical microstructure is not distinguished from that of water. This is because twater won't cause 'water' in those possible worlds in which it *is* so distinguished. A possible world in which people can make for the differences between H2O and XYZ is a world in which twater can be blocked from causing 'water'. And thus it is this difference that makes 'water' mean *water*, or express the concept WATER, rather than twater. But this should not mean that water is identical to its detection property H2O, or that twater is XYZ. This is because their chemical compositions are only a small part of what makes water and twater cause selectively the instantiations of 'water' and 'twater' in humans (or Twin-Humans). For example, we saw that as regards the differences between water and shmoter (in section 1 of this chapter), some other detection property of water becomes its "essential" property, and so on. Thus I would say that what makes the property of *being water* what it is, again from the perspective of informational semantics, is the cluster of properties that are responsible for the selective instantiations of the state of 'water' in humans, and that includes all those properties, and them only, which humans could possibly selectively detect. And correspondingly, I would say that what makes 'water' mean *water* is that it is the state which is selectively caused to instantiate by the detection properties of water, rather than, say, those of twater.

But here someone might object as follows. Who says that when the scientists determine that the chemical structures of water and twater are distinct in the way we have fantasized, that we would come to think 'water' in the presence of water rather than, indeed, in the presence of twater? After all, if all there is to our concepts is the distinguishability of the corresponding properties, then in the same way that water could cause 'water' when we do the chemistry, and therefore not twater, we can presume that twater would, whereas water would cause 'twater'. Surely, barring the actualist clause, it should make no difference what 'water' means as long as it is distinguishable from 'twater'. And a case where 'water' and 'twater' mean *twater* and *water* respectively would also satisfy that condition.

Here is my answer. First, I think the objector is right that it does not matter if 'water' means *water* or *twater*, so long as it does not mean their disjunction, as was my concern to show. But if so, then there are now two possibilities before us. One possibility is to concede the contingency of 'water' meaning *water* rather than *twater* and then make an arbitrary choice as to which of the two it would mean. Thus, I can imagine the response of our scientists to the discovery of twater on Twin-Earth in the way of tossing a coin to decide

whether 'water' should mean water or twater. But of course, there is an easier and more reasonable option which is that instead of tossing a coin, we shall just stick with the stuff *around here* and decide that *this* is water. *For if it makes no semantic difference if 'water' means water or twater, then it should make no difference if we decide that it means water.* And this, in my opinion, is all there is to the actualist clause in the semantics of natural kind (as well as other "indexical") terms. Semantically, we could have decided to make our terms mean properties to which we bear no causal-historical connection, only counterfactual ones. But pragmatically, it seems more reasonable to stick with what we know. To be sure, that is a far cry from the claim that semantics is essentially indexical (Kripke 1972, Putnam, 1975), much less the triviality that when the differences do not matter, you might just as well go with that with that which you are familiar.[197]

7. On Narrow Content and Disjunctive Properties

According to my theory, informational semantics builds on informational relations between mental tokens and the *detection* properties of horses, water, etc., rather than between token mental states and the detected properties themselves (*being a horse*, *being water*, etc.). I would like now to show how that model fares with respect to the AD and RC conditions on content as formulated by Fodor, but I will do this via a consideration of an argument which purports to show that Fodor's AD condition implies that content is in fact narrow. The reason for this comparison is that on both Fodor's theory, according to this argument, and on mine, content is disjunctive, relating to all those things that could reliably cause the instantiations of the content-bearing state. But the upshot is that according to my theory, disjunctive content is still compatible with *wide* content, or at least, this is what I would like to argue.

There is an argument in the literature due to Antony and Levine (1991) which purports to show that Fodor's AD condition actually leads to a theory of narrow content. It is called 'the proximal stimulus' argument and goes roughly as follows: Let us suppose that when horses cause 'horse's it is sometimes because they are HL. But HL can be seen as a proximal stimulus to horses, say

[197]The above considerations do not bear on the differences between the concepts HORSE and CDN. Unlike the case of water and twater where there are no nomological differences in their relation to all of their properties which might trigger 'water'/'twater', (what I called WL), cdn's relation to HL is much weaker than that of horses to HL. A world in which cows on a dark night cause 'horse' but horses do not is one in which cows (on a dark night) look like horses but horses don't. I find it hard to imagine such a world.

in the form of the retinal projection of a pattern of light, that causes 'horse's. In this case we have it that horse-caused 'horse's depend on HL caused 'horse's, for without the proximal stimulus there would not be a visual perception of horses. Whereas HL's causing 'horse's is not dependent on horses, as we saw, for cdn's are also sufficient for HL (figure 6.3).

horse → HL → 'horse' ← HL ← cdn

Figure 6.3: The dependence of 'horse' on HL rather than on either horses or cdn's.

Still, this is not sufficient to show that 'horse' *means HL*, that is, that HL is the narrow content of 'horse'. To show that, there also need to be non-HL caused 'horse's (the robustness condition). But neither horse nor cdn can satisfy this condition since any horse or cdn causing 'horse' is also an HL causing 'horse'. That is, the point is that neither horses nor cdn's can be the wild causes of 'horse's because they are not independent of HL: every horse or cdn caused-'horse' is also an HL caused 'horse'. But RC requires some cause of 'horse's which is independent of the putative veridical cause, HL in this case.[198]

So how about some other proximal stimulus, such as that relating to horse whinnying (HW)? Surely, HW can cause 'horse' without HL causing horse. Well, although this will solve the robustness problem, it will sever the dependency relation between 'horse's and HL, for now HW will do just as well. That is, it now follows that horse caused 'horse's no longer depend on HL-caused 'horses' since horses can cause 'horse's via their whinnying sound. OK, so now, how about the disjunction of *all* the proximal stimuli (abbreviated by A&L as P(INF))? Here AD is satisfied since no horse can cause 'horse' unless P(INF) does, and robustness is satisfied since a non-proximal stimulus, such as a hay-thought, can still reliably cause 'horse's.[199] But this would imply, A&L claim, that 'horse' should mean P(INF), the narrow content of 'horse', and not the property of being a horse (its wide content).

---

[198]This requires an independent *cause*, not an independent *law*.

[199]Such hay-thoughts cannot be thoughts that have arisen due to interaction with horses, since presumably this would be via some proximal stimulus. But then, hay thoughts are not independent of P(INF). Whether there *are* such thoughts is, in my view, open to question. But I won't quibble about that.

horse → P(INF) → 'horse' ← hay-thought

Figure 6.4: 'horse' means P(INF)

Fodor's objection to this argument was to the implication that open disjunctions can play the role of nomic projectible properties. I will not have the opportunity in this dissertation to defend the claim that at least *some* open disjunctions are perfectly kosher nomic properties. Still, I would like to show how my theory fares with respect to this model. To begin with, as with the dependency relation between horse and P(INF), my theory also claims that horses would not have caused 'horse's unless *either* of their detection properties did; that is, unless the disjunctive property which I will call HD(INF) (for 'Horse Detection'), causes 'horse's. And to be sure, 'horse' can be caused by HD(INF) without being caused by horses, as when it is caused by cdn. For if cdn is sufficient for HL, then it is also sufficient for HD(INF). So we now see that AD (i.e., Fodor's asymmetric dependency condition) can be satisfied with respect to HD(INF) in that the horse-caused 'horse's depend on HD(INF)-caused 'horse's, but HD(INF)-caused 'horse's do not depend on horse-caused 'horse's since cdn is also sufficient for HD(INF).

However, RC also needs to be satisfied, and the question now: is there something that is not HD(INF) that can cause 'horse's for robustness to be satisfied? Well, it seems to me that *cows* (on a dark night) satisfy that, or rather, the disjunction of the detection properties of cows, CD(INF). Although a cow on a dark night is sufficient for HL, and hence for the satisfaction of HD(INF), cdn's have their own detection properties, all those relating to *cows* plus the HL disjunct. So when a cdn causes instantiations of HL, it can do so either by itself instantiating HD(INF), or by instantiating CD(INF). If the latter, then this is a *different* cause of 'horse' than HD(INF). For although we are assuming that CD(INF) and HD(INF) overlap at the HL disjunct, they are still *distinct* disjunctions, according to my theory. What makes them distinct disjunctions are such disparity in their disjuncts, as that between horse-DNA and cow-DNA, horse-bone structure vs. cow-bone structure, and so on.

Thus, according to this model of the asymmetric control of the instantiations of the 'horse' state, 'horse' means the modal open disjunction of all those detection properties we have associated with horses, HD(INF), rather than with horse itself. And by parity of reasoning, the same goes for all other concepts, whether they are of other natural kinds, of artificial kinds, or of no kind at all.

8. Summary: The Definition of Mental Content

It is time to wrap the discussion and give my definition of mental content. I say that a state 'x' of an organism *s* means *X* iff

1) The Information Condition: 'x's carry information about X's or, "x's iff Xs' is a *ceteris paribus* law.

2) The Robustness Condition (RC): There is a Y (which is not an X) such that Y→ 'x' is a *ceteris paribus* causal law.

3) The Asymmetric Dependency Condition (AD): The causal law Y→ 'x' is asymmetrically dependent on the causal law X→ 'x'.

All of the above conditions are ones which have been in the literature in some form or other. My particular contribution is in the following lemma:

Lemma: For all Y (which is not an X), there is a cognitively possible world W such that in W X's nomically cause 'x's in humans but it is not the case that any Y's do.

Let's call the lemma the Modal Response-Dependent Supplementary Conditional (MRDSC; read 'mardsack'). MRDSC is not an extra condition on the definition of mental content but rather a supplement to the AD condition in that it shows how the wild causes of 'x's, Y's, depend on its veridical cause X, but not the other way around. The asymmetry is shown in that MRDSC entails that there is at least one possible world in which only X's cause 'x's, but does not entail that there is a possible world in which only Y's nomically cause 'x's. To verify that this is indeed the case, let me relate to three possible objections that I have heard against it.

Objection One: Suppose there is a possible world W* in which *it just so happens* that there are no X's but there are some Y's which cause 'x's.

Objection Two: There is a possible world W* in which there are both X's and Y's, but *it just so happens* that by some cosmic arrangement of light Y's look like X's and X's look like Y's and hence only Y's cause 'x's.

Objection Three: Like Two, but here there is a mad scientist who makes it such that Y's look like X's and X's look like Y's and therefore only Y's cause 'x's.

As we can see, all the objections stipulate the 'it just so happens' clause in their formulation of the aberrant worlds. Therefore all the objections presuppose in their very formulation presuppose a violation of the assumption that the content constitutive relations must be nomic relations, and that the worlds must be cognitively possible worlds, that is, worlds in which the laws of psychology hold.

Thus, objection One ignores the fact that the relation of X's to 'x's is presumed to be a nomic informational relation which holds even in worlds where there are no X's. It follows that even if there is a world W* in which there are no X's and such that only Y's cause 'x's, still, then surely *if there were* X's in W*, then they would also cause 'x's. But this is not the case with respect to W. Here, even if there were Y's, only X's would cause 'x's. So W, but not W*, is a world where Y's *are* distinguished from X's.

Objection Two: Here it is assumed that worlds in which such cosmic illusions occur are still cognitively possible. But I don't believe so. This is because I think that our psychological discriminatory capacities have developed under certain environmental conditions, those which obey the laws of nature. Hence, an environment which is not itself nomically possible is also not psychologically possible. So W* cannot satisfy MRDSC since the lemma applies only to psychologically possible worlds.

Objection Three is somewhat different as there is no psychological law which is violated: mad scientists are certainly a psychological possibility. But for this very reason it just begs the question. In order that the mad scientist be able to pull the stunt of making Y's look like X's and vice-versa, she or he would already have to possess the relevant concepts of X and Y. This means that there are humans (even if only *possible* humans, if the thought- experiment is about a *possible* mad scientist) who can distinguish X's from Y's, which implies that MRDSC is already satisfied.

But suppose that the mad scientist is *not* human, say a Martian scientist, who came to visit earth. In fact, we can even suppose that she is a benevolent scientist who wants to help us into further discriminations in nature (say, between water and *imp*water). Would that possibility count as a W world? The question here is, I believe, about the nature of the covariance mechanisms that we are willing to allow between our mental states and properties in the world. Can an informational law be implemented by a mechanism which is not psychologically possible for humans, although it is for other "higher" creatures? (We do not have to assume that we need to *communicate* with these creatures). If the answer is 'yes', then we would also have to allow for some worlds where the mischievous Martians fiddle with our access to the world. Well again, the answer is 'no', since I don't think that Martians are a psychological possibility *for us,* in the same way that wild cosmic rays were not a psychological possibility *for us.* Had we developed in a world in which Martians play a significant role in our environmental interactions, we could have factored the Martian element into the very structure of our psychological make up. Things being as they are, Martians are out: we have to rely on ourselves, for better (no W* worlds with vicious Martian scientists) or worse (no W worlds with benevolent ones).

There are still a few issues left open with respect to my informational theory. First there is the issue of how my theory fares with respect to the presumed problems of necessary coextension and necessary coinstantiation, problems that according to some are the most serious ones for informational semantics (see for example Egan 1991). Then there is the opacity problem for content, and the question about the scientific legitimacy of disjunctive properties. In particular, that concerning the possibility that a disjunctive property like HD(INF) can play the role I assigned to it in the determination of content, as a nomically kosher property. Unfortunately, the full development of these issues will have to wait for another occasion. In the next chapter I will summarize the theory of informational content I have developed in previous chapters and analyze where I think it stands with respect to the desiderata I have outlined in the introduction. I will close the dissertation by saying a little about the theory of properties which I believe has to be presupposed by my informational theory of mental content.

Chapter Seven

The Modal Response Dependence Informational Theory

1. Précis of Previous Chapters

As said, in this chapter I intend to bring to completion my thesis as regards the theory of intentional-mental content that I have advocated at various points in previous chapters. Part of the burden of this rundown will be also to show how I think this theory fares with respect to some of the problems for mental content that I have reviewed on the way, especially the desiderata listed in the introduction. I therefore believe that a summary of what transpired in chapters 1-6 will be expedient here.

Chapter One dealt with the role of intentional content for mental causation within the framework of an intentional-nomological conception of psychological explanations and predictions (PEPs). According to this thesis, mentalistic explanations are a species of causal-nomological explanations used more generally in the natural sciences. The background assumption is that given a minimal amount of scientific realism, confirmed intentional causal laws should point to an underlying network of causal interactions between events subsumed under intentional descriptions. In this way, proponents of the intentional-nomological conception of PEPs try to justify their claim that mental events cause and are caused *qua* mental events.

The idea that the causal role of mental states is constrained by their intentional properties was expressed by Jerry Fodor as the thesis of isomorphism. This is the claim that there is a structural similarity between the nexuses of mental processes, on the one hand, and the semantic relations that obtain between their propositional objects, i.e., their contents, on the other hand. I called this tenet of intentional realism the isomorphism thesis, or IT. However, we saw that IT seems to violate Hume's stricture that causes should be logically independent from their effects, something which was used by certain philosophers (the 'anti-causalists') to argue against the possibility of intentional causation more generally.

The challenge in chapter One was then to reconcile the causal aspect of mental processes with the semantic constraints that *ex hypothesi* govern or control them. My solution was to present a complex dependency relation between the two domains, the causal and the intentional, via the model of Cummins' theory of property instantiations. The idea is that a physical process, though implementing an intentional process, is not *sufficient* for it, and hence not *dependent* on it (that is since sufficiency creates dependency: when B is sufficient

for A then A is necessary for B). The intentional process, on the other hand, while dependent on there being *some* physical process, is not dependent on the particular one which actually *realizes* it. How do we make a physical state or event realize one intentional property in one situation (or possible world) but not in another? My claim was that this had to do with the peculiar nature of intentional properties that they are constructed from open disjunctions of physical properties, such that the same disjunct, say, a certain movement of the hand- muscles, can figure in different disjunctions. In any particular case where such a disjunct is instantiated, this could be because of the overall instantiation of one disjunctive property or another: either because of the instantiation of the disjunction of those possible physical behaviors associated with wanting to send your fellow climber to a better world, or due to the instantiation of all those possible ways to become erratic. More than this, I also claimed that for some physical state to count as realizing an intentional state, it is *necessary* that in some possible worlds it is instantiated as part of a different disjunction which does not realize that intentional property. (This is in line with Fodor's asymmetric dependency condition for content that I discuss in chapter Six). I then claimed that the decision as to which disjunctive property is actually instantiated is a matter of evaluating some relevant counterfactuals (e.g., what would the climber do if the rope got stuck in a crack in the rock, etc.). The metaphysical background to this was then presented as one concerning the topology of the logical space of property relations.

Let us suppose that our climber is the vicious climber. He had already a few similar instances in his past, and in many of the nearest possible worlds to this one he is making sure his fellow climber lets go of the rope. In which case, we can be sure that in all those nearby worlds this very same physical disjunct concerning the hand movement is adjoined to the other disjuncts which together make it a vicious act ('cutting the rope *or* shaking it back and forth, *or....' etc.*). Alternatively, if he is the decomposed climber, then the physical responses would consist of the likes of excessive sweating *or* twitching, *etc.*). The point is that it is the topology of this logical space and its distance metrics[200] which will eventually decide the issue of which disjunctive property was instantiated in the actual world, again, depending on the relevant counterfactuals.

In light of this solution to the problem of whether reasons can be causes, and given other considerations raised in chapter One (for example, that the problem of "explanatory overdetermination" is not unique to mentalistic explanations) I conclude that there is no hindrance to construe content as a nomic property of mental states. And given Davidson's own argument that

---

[200] On the idea of possible-worlds metrics see Boghossian 1991, and Fodor 1991, ch. 4.

reasons must be causes so that mentalistic explanations have an explanatory *force*, I conclude further that content is a causal-nomic property, and thus that desideratum A is satisfied.

In chapter Two the isomorphism thesis was at the center of the discussion. The thought was to substantiate IT via the verificationist theory that the content of the attitudes consists in the inferential relations between their propositional objects, and connect it to the functionalist theory of mental states: the idea that such states are defined by their causal roles. We saw that just such a theory was offered under the title 'conceptual role semantics', or CRT. The rest of the chapter was dedicated in the main to a myriad of arguments for and against CRT. Three main problems stood out in particular: the problem of holism or meaning incomparability, the problem of the representational aspect of content, and the problem of "grain" of content, the latter posing special difficulties for the empirical adequacy of inferential role semantics. I then concluded the chapter by saying that to avoid these problems, we need to look for a notion of content which is atomistic, representational, and with a level of "grain" that the data supports. I claimed that we could achieve all these desiderata if we construe content by reference to laws which connect intentional states with fine-grained properties in the world. I then insinuated that informational/truth-conditional theory is just the theory we were looking for. Informational content is atomistic in that informational, or covariance, relations to the instantiation of properties in the world make no essential reference to other concepts. It is representational in that it is *about* objects and states of affairs in the world, and its grain can be determined by the level of grain of the properties with whose instantiation information-bearing states covary.

Thus, in chapters Three and Four I took a holiday from the issues of mental causation proper in order to focus on the task of finding a semantic notion of content which would be both informational and empirically adequate for psychological theorizing. In chapter Three I claimed that the kind of content which would best suit our needs has to be informational like Russellian propositions, but with the level of grain of Fregean senses. I suggested that we combine the two theories to concoct the right notion of informational, finely-grained content. From Russell to take the idea that meanings are the things and properties in the world, and from Frege to take the idea that coreferential/coextensive sentences and terms can express different information content. 'Marrying' thus Russell's theory with that of Frege gave rise to the theory of content in terms of the worldly instantiations of fine-grained properties and truth-conditions. I concluded that although it might be said that (e.g.), the proper names 'Cicero' and 'Tully' are *about* the same individual, they *express* different properties, that of *being Cicero* and that of *being Tully*. Although the two properties are coinstantiated, they are still distinct properties. The idea was that if this semantic conclusion was to have the right bearing on the question of

mental content, then, to continue the example, Cicero thoughts would be considered distinct from Tully thoughts, something which could explain why someone can believe that Cicero was a nice guy but Tully was a naughty fellow.

Chapter Four continued the line pursued in chapter Three by concentrating more directly on the question of mental content in the context of the relational theory of the attitudes: the idea that mental states involve a relation between an agent and a proposition. At the beginning of the chapter we encountered a classical argument, the slingshot, that purported to show that all the grain gained by adopting Fregean propositions as the objects of the attitudes would be lost if we accept that sentences, or 'that'-clauses, can serve as *names* for propositions. Indeed, at this point I claimed that an alternative to the relational theory exists in which 'that'-clauses express representational properties. But I also noted that the slingshot rests on other questionable assumptions, one being that all logically true sentences express the same proposition. And of course, this cannot be accepted by a theory which aspires to come-up with a notion of content that distinguishes the meanings of logical truths. After all, it was my conclusion from chapter Two that verificationists went astray partly because they took equivalent propositional attitudes (for them, those which licence the same inferences) to be identical in content, *pace* our empirical findings.

The situation at this point with respect to the search for the holy grail of mental content was as follows. I turned back to Frege and Russell to retrieve a notion of content which would be as fine-grained as Fregean modes of presentation, and as informational as Russellian propositions. But that attempt hit bedrock with necessary truths. For necessary truths are truths in all possible worlds, which means that they all have the same truth-conditions. Even if we think that there is a sense (*pace* Kripke) that Cicero might not have been Tully (after all, that *sounds* possible), what would it mean to say that a triangle might not have been a trilateral, or that a bachelor might not have been an unmarried male? I had therefore to continue the search. I took a look at Bealer's theory of PRPs, which promised to be a theory which could pull the trick and hyper-fine-grain content, but then we saw that Bealer's theory relied on a vague and in fact implausible distinction between simple and complex concepts. Next, I reviewed some meta-linguistic suggestions, but for a theory which is looking for a purely informational notion of content, this was much work for nothing. At most, meta-linguistic differences can convey information about linguistic use, not about the way the world is.

At this point in the dialectical order, the author of the dissertation woke up from his dogmatic slumber and cried out: 'Of course we cannot find a fine-grained enough notion of propositional content, one which is more fine-grained than all possible inferential relations. This is because propositions are, *by definition*, just those operations closed under the logical (or inferential). I needed to continue the search, but this time to change the strategy. Instead of looking for

a *propositional* notion of content by digging into all sorts of semantic theories - all the while keeping one eye open on the constraints coming from our psychological theorizing - why not dig for that notion of content in the mind itself? Thus, my suggestion was to engage in what I called 'the method of Psychoanalytic Ontology'. The premise from which this idea proceeds is that we can get the best inkling about the identity conditions for contents by knowing more about what *concepts* there are in humans. Thus, instead of adopting the slogan 'to the things', we should rather go with 'to the concepts', that is, the meanings in human minds, and do this in the way of constructing psychological theories in which care is taken to sort out the concepts people have from the noise that accompanies them (e.g., knowledge/ignorance of linguistic conventions, use of distinct mechanisms, etc.). Indeed, the idea here is just that of using human systems as indicators, more or less reliable, but possibly the *only* ones available, for the instantiation of concepts.[201] But we should observe that with all that, I still haven't deserted the idea that concepts correspond to the instantiations of fine-grained properties in the world. Quite the opposite: it is only on the assumption that informational connections to properties in the world constitute the conditions on content, that the method of 'psychoanalytic ontology' can make sense in the first place. Only when we suppose external truth that it makes sense to sort out genuine concepts from those contaminated by linguistic, psychological, and other collateral information.

Thus in the method of 'psychoanalytic ontology' I have found at last a way to satisfy desideratum D1 which requires a notion of content with the level of grain that is empirically adequate; a notion of content which reflects the functional role of concepts in humans psychology. But unlike CRT, it is not claimed that such functional roles are *determinative* of contents, only that they are indicative of content. The determination relations, consonant with the informational theory, are with the relevant instantiation of properties in the world.

With these temporary achievements in hand, I returned in chapter Five to the issue of mental causation, this time with respect to the position which Burge (1979) dubbed as individualism. The question which loomed large there was how to reconcile an informational notion of content - wide content - with the classical conception of causation as a local affair. For the suspicion raised by opponents of a relational individuation of mental content was that wide content reintroduces the problem of epiphenomenalism I touched upon in chapter One. Once again the question came about how mental states can cause and be caused *as* mental states, although the context has somewhat changed.

---

[201] I should note that this process is no more individualistic than any other inductive method of arriving at generalizations via the study of particulars.

In my presentation of the topic of individualism, I stressed that there are in fact *two* individualistic theses. There is Burge's individualistic thesis according to which mental states should be individuated by reference to the *explanatory* constraints from FP or from science. Then there is Fodor's individualistic thesis according to which mental states should be individuated by reference to their causal powers. As we saw, neither thesis entailed whether mental states should be individuated by reference to the intrinsic or extrinsic physical properties of individuals. Still, Burge has claimed that at least some psychological explanations make reference to wide individuation schemes, and hence that wide content psychology cannot be ruled out a priori. Fodor, on the other hand, has claimed that whether psychology can rely on a notion of wide content or whether it needs instead to adopt a revisionary notion of content, narrow content, depends on whether wide individuation of mental states is done by causal powers or by what affects causal powers.

Now although Fodor argued against relationally individuated mental states that they are neither individuated by causal powers, nor by what affects causal powers (two formulations of the constraint on individuation which, I argued, amount to the same thing), we in fact saw that the real issue here was the supervenience thesis. For as long as the supervenience of the mental on the physical is satisfied, one can transfer questions about individuation by causal powers from the realm of mental properties to the realm of their supervenience base. But mental properties individuated widely can be supervenient on physical properties only if brain states can themselves be individuated widely.

So then, when I turned to the question whether there is a wide individuation scheme of brain states which also individuates them by causal powers, my claim was that there is, if indeed we adopt an informational theory of content. For informational relations are constituted by covariance relations between states of the organisms and features of the environment, and their underlying mechanisms can endow brain states with relevant causal powers. A brain state which is tuned to instantiations of water in its vicinity has surely different causal powers than one tuned to instantiations of beer, left shoes, and even twater (given the right counterfactuals). We just have to agree that the *kind* of brain states which acquire these causal powers probably cross-classifies with neural kinds which have their non-mental, neural functions in the brain.[202] Thus at one stroke we both allow a wide individuation scheme of intentional mental

[202]Either across individuals or even in the same individual. I would even go as far as to say that it might be possible that the same neural kind will have one token of which carries some physiological work, and another token of which carries environmental information. This is what I would call a true division of labor.

states as truth-conditional semantics has it, and as Folk Psychology has it, and at the same time we preserve the metaphysical thesis of the supervenience base of causation, namely that the causal powers of mental properties should supervene on those of physical properties (premise B in Fodor's argument presented in chapter Five). In this way we are able to maintain the basic intuition behind individualism concerning the supervenience thesis, without subscribing to the MS thesis that mental states need to be individuated by bracketing (i.e., abstracting from) their semantic properties.

From the above we can predict what implications this model will have for a science of the mind. On the one hand we see that cognitive psychology, as a psychology sensitive to the informational content of brain states, would be "in", while neurology, which is the science that studies brain states as neurological kinds would be "out". As to computational psychology, the position advanced by Pylyshyn and Fodor (and of course others): since computational processes are assumed to realize informational processes, it would make sense to assume that computational states are individuated along the same seam lines as informational states are. In other words, computational states are just one more way to individuate brain states which are implicated in information processing mechanisms. In fact, computational states form a major part in those mechanisms, as Marr's theory of early vision exemplifies (Marr 1982).[203]

My claim that widely individuated mental states are in fact brain states which bear certain informational relations to the environment also helps to sort out a certain puzzlement concerning wide mental individuation, as well as relational individuations more generally. We saw that one of the major obstacles that externalist theories were facing was to account for the explanatory power of relational properties. According to Fodor, a relational property becomes an explanatory property once it affects causal powers. In this way, Fodor attempted to distinguish genuine explanatory properties from those involved in mere 'Cambridge' changes (see chapter Five section 4). Although this is an "a priori criterion", it can also be used in actual explanatory practices, as when we attempt to explain why such a property as *being a planet* is an explanatory property. The

[203]In this respect, there is no conflict between Burge's claim that Marr's theory is intentional (in Burge 1986 part II) - in that it involves reference to the environmental features implicated in what Marr called "the computed function" (Marr, ibid.) - and the claim of computationalists that the theory should be deployed at the algorithmic level. Indeed, Marr himself thought that his theory can be deployed at *three* levels of description, including also the "hardware implementation" level. In my view, all these are compatible ways of individuating a class of brain states which are engaged in the same information processing tasks.

answer is that the relational individuation of chunks of matter is explanatory because *being a planet* affects the causal powers of certain chunks of matter, probably by affects their kinetic energy. In contrast, it is hard to find in the writings of externalists an explanation for why the relational properties that they have sampled from the sciences are explanatory, mostly just accepting that they are.[204] Now surely accepting the authority of a science is a well respected methodology, but one would expect that a philosophy of science would also try to go beyond that and explain the practices that the sciences engage in, that is, explain the explanation.[205] I have claimed that the idea of informational relations and informational mechanisms underlying relational individuations in general could disperse the mystery. But whether this is true or not in the other sciences, it follows as the most reasonable supposition for psychology. Since my view is

---

[204]Wilson, for example, asks us to consider the relational explanation of the movement of two particles identical in all their intrinsic physical properties by reference to the differences in their surrounding magnetic fields. He says: "there is nothing causally mysterious in saying that [these particles] move with different velocities *because* of their locations in different magnetic fields, even though being so located does not supervene on the intrinsic physical properties that each instantiates." (1995:143)

However, as Wilson seems to acknowledge, this is not an example where there are *no* local physical changes (in the charge of the particles) and hence not an example where we have *no* explanation for why the relational explanation works as well as it does. As Fodor would probably say, here the magnetic field *affects* the causal powers of the particles, and so there is indeed nothing mysterious in this. Indeed it seems that Wilson *does* accept the need for a mechanism to make for the causal efficacy of relational properties, and in that sense he is not a Burgean externalist, as I have claimed. For even though he says that "There is no special type of causal relation or causal mechanism required to make sense of wide psychological differences between individuals... even doppelgängers" (ibid., p. 144), that is not because *no* mechanism is required, but because "causal mechanisms can bring about different causal effects because of differences in their historical or relational properties". (ibid.) By the way, I agree with "relational", I disagree with "historical": Mechanisms are essentially counterfactual.

[205]Externalists can of course take the Wittgensteinian line and argue that 'explanations have to stop somewhere', thus excusing themselves from the need to explain the explanations. This is their right. Still, I believe that some people would prefer a theory that answers that question to one which merely dodges it.

that mental content is informational content, it is a short way from this to give a mechanistic explanation of the explanatory role of relational mentalistic properties in the science of the mind.

With this I can now conclude that informational theory of content has the resources to satisfy desideratum D2 as well. This was the condition that a theory of Mental Content be able to explain in effect why not every relational individuation of a psychological state is individuation by content. We now see that although informational theory is 'externalist' in that it individuates states by reference to some of their external relations, this is true only when there is an informational mechanism which underlies these relations. Context, according to this theory, makes for scientific taxonomy only when it is anchored, in some nomic way, to states of the individual. Just being 'embedded' in a context cannot be enough for individuation. On the other hand, since the class of brain states that form informational connections to worldly affairs can cross-classify with neural kinds, molecular Twins can still instantiate states with distinct informational content, as desideratum requires. Thus, Putnam and Burge's intuitions about environmental individuations (but not social ones) can be satisfied, although their explanation for this phenomenon is not the one informational theory gives. According to informational theory, it is not the causal-historical relations which matter, but the counterfactual ones. Both Oscar and Twin-Oscar can have water-thoughts (or twater-thoughts), even if they come with different histories of linguistic or conceptual acquisitions. If there is no possible world in which they can distinguish water from twater (with the use of experts etc.), then water and twater mean the same for them both: it is that stuff whose instantiation would selectively cause them to think WATER. My claim was that since for both *there is* a possible world in which they can distinguish water from twater, the content of WATER in their head can be either *water* or *twater*(although not their disjunction). But in fact I suggested that pragmatics dictates that the content of their thoughts in such a case would be determined by what is available (or salient) in their vicinity.

2. The Informational Theory of Content

After quite some build-up, in chapter Six I arrived at last to the theory of informational content itself. I first defined informational relations in terms of causal-covariance relations, or laws, between states of the system and worldly events. But then we saw that not every such covariance relation can constitute a semantic or a representational relation, since the latter also require the possibility of error (desideratum G). While informational relations are always veridical, semantic relations tend at times to give rise to misrepresentation (this is why we have 'pan-informationalism' but not 'pan-semanticism'). The problem of error in the context of a causal theory of content gave rise to the disjunction

problem. On the one hand we require that an informational state could be reliably caused by events which are not in its extension. Thus we require that not only X's would cause 'x's but that Y's, distinct from X, could cause them as well. But on the other hand, a state will carry information about anything which nomically covaries with it, in this case (X v Y).

To solve the disjunction problem, Fodor suggested the asymmetric dependency condition. According to this, the process going from the wild cause(the Y's) to the instantiation of the content bearing state would be metaphysically dependent on the process going from the veridical cause. The idea is that every possible world in which the wild process is taking place is also one in which the veridical one is taking place, but the veridical process can happen on its own. But then we saw that AD suggested that there is a common mechanism underlying both processes. My reasoning was that if the wild tokenings are dependent on the veridical tokenings, then this is probably because the mechanism which underlies the instantiation of the token bearing state is sensitive to features common to both causes. For example, if both cows on a dark night and horses reliably cause 'horse's under similar conditions, then this surely is because of something that horses share with cdn's, and to which the underlying mechanism is sensitive. Similarly, if both water and twater reliably cause 'water' or cause 'twater', then this is because of features that water and twater *ex hypothesi* share.

At this point it seemed that we have returned back to the theory of narrow content. This for the reason that, at least on the face of it, the features that water and twater share, and those that horses and cdn's share, are their *manifest* or *phenomenal* properties - those features which cannot distinguish the physical responses of molecularly identical Twin. Another problem was that if the mechanism that mediates the response to the veridical causes was the same as that mediating the response to the wild ones, it is hard to see how the satisfaction of the asymmetric dependence condition could even come about. For as I have just noted, the asymmetric dependency condition requires worlds in which only the veridical cause would trigger the activation of the content bearing state, even in the presence of the wild instances. But as the mechanism is sensitive to features common to both wild and veridical instances, there can be no such difference. It therefore seemed that an informational theory of content collapses into a narrow theory of content.

My solution was to claim that the possible worlds in which wild instances fail to cause instantiations of the content bearing states while the veridical instances do, are worlds in which the wild instances fail to instantiate some of the properties to which the mechanism is sensitive. For example, I claimed that a world in which horses cause 'horse' but cdn's do not is a world in which cdn's fail to instantiate horse-lookingness. This is of course a possibility since cdn's are in fact cows (-on a dark night), and cows are not essentially horse-looking. Thus

one can quite easily make cows look like cows rather than like horses if one changes their relation to the property of horse-lookingness, say by using a flashlight, or a night-vison equipment, etc. Presumably, this is easy to do since, *ex hypothesi*, the only common thing that cdn's bear to horses is their horse-lookingness they acquire under certain illumination conditions. The case is somewhat different with water and twater, which share not just one property but, again *ex hypothesi*, *all* but one. Still, because of this one difference, twater can be separated from its contingent property of water-lookingness, or more broadly, water behavior, by the use of more 'sophisticated' tools which can detect its chemical microstructure. Of course, this should not mean that for 'water' to mean *water* rather than *twater* this chemical difference has to be manifested in every world in which water (or water and twater) causes 'water'. Rather, because informational theory is a counterfactual theory, it is enough that there is at least *one* cognitively possible world in which this takes place (and ditto for the case of horses and cdn's).

The theory of content which arises out of this model is thus the following. A state 'x' of the organism means that property whose instantiations, under certain conditions, would exclusively and nomically cause 'x's instantiations in at least some cognitively possible worlds, given that there are also cognitively possible worlds in which non-X's cause 'x'. Or to put it more formally:

MC: for all X, 'x' means X iff 1) X→'x' is a causal-informational law, 2) there is a cognitively possible world W* in which Y's (which are not X's) cause 'x' and 3) there is a cognitively possible world W in which X's can be distinguished from Y's such that only X's cause 'x's in W.

The central concept in MC, as I see it, is that of the 'cognitively possible'. The idea here is to constrain the set of possible worlds quantified over in the definition to those in which human cognitive systems can make the relevant distinctions. And those worlds, it should be noted, form a quite heterogenous bunch. On the one hand we have worlds where the laws of perceptual mechanisms hold, which form a small subset of the laws of nature. On the other hand we have those worlds in which humans have developed their "ideal" theories of nature, which might just be worlds which are coextensive with the physically possible. But however we organize this set of possible worlds, the point is that our discriminative capacities with respect to the instantiation of properties around us are limited to it. Any possible world which transcends this set would be one in which certain distinction are going to remain semantically meaningless to us. Or to put it a bit differently, any distinctions which transcend all our nomologically possible discerning capacities as a species are not semantically meaningful distinctions for us.

In the last chapter I gave a theoretical example of such a property which humans cannot possibly detect, 'impwater': a stuff which is just like water but which is distinguished from it in something that is not possibly detected by humans. We can now say that this is a property to which we are 'cognitively closed', to use McGinn's term, in the sense that we are "built not to grasp" it (McGinn 1989). And of course, there is an indefinite number of such properties that are not possibly detected selectively by humans in a similar way.

Thus the first metaphysical implication coming out of my version of informational semantics is that there are certain truths which we would never comprehend or understand, if we assume that such capacities are constrained by our cognitive resources. In this sense, my theory is 'anthropocentric', although it is consistent with a position of metaphysical realism, perhaps some will even say *extreme* metaphysical realism (Edgington 1985), which I take to be as common-sensical as any. To think otherwise, that is, to think that there is no truth however remote and minute which we could not "in principle" discover, as the logical positivists claimed (Ayer 1946:38ff), requires either a lot of confidence in our cognitive capacities, or a lot of pessimism with respect to the compass of truth.

There is a respect in which my theory is anthropocentric in that it takes as meaningless propositions which describe differences humans cannot possibly discern. However, because of this very fact, unlike other anthropocentric positions (Quine 1960, Davidson 1980,Goodman 1978, Putnam 1981, Rorty 1980) it is not claimed that there is a substantive set of propositions about which it is impossible for the human species to be in error or ignorance. It only implies that it is possible for there to be a set of propositions about which the human species is *not* in error or ignorance, which is a different matter altogether. The first case implies that those propositions are known to be true in every possible world, which for a large class of propositions seems to be absurd. The second case, my theory, only requires that there is some possible world in which we know those propositions to be true, which is not only a reasonable claim but a condition, I would argue, to engage in an intellectual inquiry in the first place.[206]

For example, a 'response-dependent' theory of concepts, such as that of Johnston (1993) stipulates that secondary quality concepts are dispositions for certain sensations: Smoothness is a disposition for feeling smooth to the touch, redness is a disposition to appear red to the eye and so on: Each concept and its associated disposition for sensation. From which it follows that once the relevant contextual conditions have been fulfilled, there is no possibility for error or ignorance since it is true by definition that something (e.g.) appears red iff it is

[206]This is actually sometimes cited as the motivation behind realism, the claim that we need to leave room for discovery (see Pettit 1991:593).

red. That is, it follows that if it is red then it will look red to normal observers under normal conditions (hence no ignorance), and that if it appears red under the same conditions then it is red (no error).

Although I would consider my theory as a species of response-dispositional theories (the term is due to Johnston ibid.), it differs from Johnston's version in that it allows error, as required by Fodor's robustness condition, as well as ignorance. Error is made possible when the content bearing state is caused to instantiate by a different disjunctive property than the content property, as when it is caused by cdn rather than a horse, or when it is caused by twater rather than by water. And ignorance is allowed in a similar way, since the presence of horses or water does not mean that the state instantiated will be 'horse', or 'water' rather than 'cdn' or 'twater' respectively. We also saw how error can happen in the case of intentional action, when the same physical behavior caused by an intentional state was part of a different disjunctive property than the one intended.[207]

Another theory which might be compared to my theory is verificationism, or perhaps even more accurately, the theory of Idealism which I have discussed in chapter Two (section 6). This is the theory where meanings are defined as verification under ideal conditions, including the possession of ideal instruments and theories. On the face of it, there is an idealist element in my theory since it requires that we be able to affect discriminations between the instantiation of properties at least in some possible worlds, that is, those in which we posses an ideal theory. However, there are a number of reasons why my theory is not a verificationist theory. First, on both verificationism and Idealism, content is *identified* with the methods of verification. This, as I have said, leads to holism and meaning incomparability, as well as to the problem of fusion. In contrast, in my theory, the methods and tools used to affect the selective detection of properties are just that, tools, and hence replaceable in principle. Rather, what matters for the determination of content in my theory are the properties with which the mind covaries, with the discriminative aid of the ideal theories and tools. But content is still informational/truth-conditional, even if it is restricted by those theories and tools. Perhaps the best way to stress the point is by saying that even when we fall short of discrimination, as when we cannot distinguish water from impwater, it is still the world we detect, although not as finely as we would wish. But this still does not make our actual concepts become identified with the limits of enquiry or intelligibility.

The second reason my theory is not verificationist is that it does not require knowledge of verification conditions by anyone, anytime, as a condition for the possession of a concept. The only thing that it requires is selective

[207]Here the "intended" was in fact the vicious act.

response on the part of the organism to the content property in at least some possible worlds, whether the agents are in possession of an application criterion for the concept or not. In fact, the agent does not need to be in possession of *any* concept other than the one she discriminates, so long as she stands in the appropriate discriminatory covariance relation to the relevant property. This indeed shows how that theory can be truly *atomistic* (Again, desideratum C).

The third reason to distinguish my theory from verificationism concerns statements containing unrestricted universal quantifiers, as the one given by Dummett (1978:16) : 'A city will never be built here'. Dummett thinks that a truth-conditional theory of meaning (such as that of Davidson which he specifically has in mind) cannot supply an account of an understanding of that sentence since it cannot supply a conception of "how it might be known to be true" (Dummett 1976:100). According to Dummett, the truth-conditions of the sentence are "recognition transcendent", meaning by this, I presume, that it would be impossible for anyone to grasp the infinite number of propositions that constitute its truth conditions (and their truth-values, if we also want to know if it is true). But my informational theory has no problem with such an obstacle since it does not require anyone to grasp an infinite number of propositions to know the meaning of this sentence or any other. What it does require is the notion of indefinite possibilities, which the theory of possible worlds already has, and then stipulate the following: that for any future time for which the sentence 'a city is not built here right now', there would be a possible individual who can grasp this sentence's truth.

As a result of the above discussion, I would therefore suggest that we take the informational account of content I'm trying to promote as advancing a discriminative notion of content with the modal force that has been explained above. This idea, of basing a notion of content in the discriminative abilities of agents is of course not new, although I believe that my version of it in terms of *all* cognitive possibilities has not been widely entertained. Thus, Peacocke has already suggested the 'principle of discrimination' according to which

> For each content a thinker may judge, there is an adequately individuating account of what makes it the case that he is judging that content rather than any another (1988:468).

Where by an adequate individuating account Peacocke means "a condition which distinguishes judging that content from judging any other content" (ibid.). In his justification of the principle Peacocke says that

> it is incorrect to attribute to a thinker propositional attitudes or other psychological relations to finely-sliced things if the abilities possessed by

> the thinker which might be cited to justify such an attribution do not slice that finely (ibid.).

However Peacocke seems to imply that it is *actual* discriminative abilities that the relevant justification requires, not possible ones as well. In any case, I believe that my theory would satisfy Peacocke's condition since, while I admit that I don't know exactly what Peacocke would accept as an adequate individuating account, I think that my informational theory certainly goes a long way in that direction in its capacity to supply detailed accounts of actual information processing devices. These can range from such contrived devices as thermometers and photoelectric cells to the more complicated structures of human vision as described by Marr's theory. Since we might presume that the basic structure of all informational mechanisms is the same, we can conclude that such accounts can be adequate in principle, even if not in detail, with respect also to conceptual discriminations that are only possible from our present perspective. According to the informational theory then, an adequate individuating account is one where there is a flow of information from the relevant property to the detection mechanism to the exclusion of any other property, as would be explained by an ideal theory as an extension of current operating theories. It is this sort of an inductive account that can enable us to quantify over theories and other instruments in the definition of our concepts.

3. The Theory of Modal Disjunctive Properties (II)

So far my discussion of the cognitive capacities that underlie the modal human responses to properties has concentrated on the receiver side, that is on the organism. No less important in this theory is the discussion of the properties which cause those selective responses. From the perspective of the informational theory, the right way to describe those properties is as disjunctive properties, i.e., as a disjunction of all those detection (or 'access' as I called them in chapter One) properties that would cause selective responses in the organism. In some possible worlds those selective responses would be measured with respect to one or a few other competing properties, as in the case of water versus vodka, or water versus beer, etc., but in other possible worlds they would be measured against all competing properties. But since the capacity of such a property to affect a selective response need not be manifested completely in one possible world, it can be viewed as a disjunctive capacity. In might be claimed that what is disjunctive about those properties needs not be in their own internal structure but rather in the structure of modal response on the part of the relevant

intentional system.[208] However, because I have already claimed that we are cognitively closed with respect to the internal structure of those properties, I will continue to assume the "internal" perspective of information theory and treat those properties as disjunctive themselves. Hence the Kantian structure of my informational theory.

According to the metaphysical components of my informational theory of content, content properties are constructs out of disjunctive properties over clusters of detection properties on their modal behavior in causing the instantiations of the corresponding concepts. Thus, for example, the cluster associated with water would cause the selective instantiations of 'water' not all in the same world but overall: in all those possible worlds in which water can be distinguished from non-water (where I have assumed the limiting case to be something like twater, or shmoter). This modal behavior of water and other content properties, which was expressed by MRDSC as quantification over a set of cognitively possible worlds, is translated by the metaphysical theory into a 'modal disjunctive structure' of those properties, depicted graphically in figure 7.1. In fact, this modal structure is exhibited in two ways in the structure of the content properties. One way is in the *quantitative* differences between content properties, that is, by reference to the number of detection properties in which they differ; for example, when horses differ from cdn's in all but the HL position. This quantitative difference goes back to the quantitative definition of information content discussed in section 2 of chapter Six (p. 175) where the amount of information at the receiver is measured by reference to the reduction in the number of possibilities at the source. But we can now see that in effect there is no way to measure the *amount* of information that mental contents register when they are instantiated since the number of possibilities of potential causes of 'x' is indefinite. Hence, the amount of information registered by mental content is also indefinite. I think this result squares with our intuition that content cannot be expressed quantitatively. Which leaves us with the qualitative aspect of mental content (which we should distinguish from the phenomenal aspect).

The qualitative aspect of content is expressed in the qualitative differences that exist between the ways to affect discriminations between properties, which goes in turn to the "strength" with which entities instantiate their detection properties. As an example, I have in mind the strength with which horses instantiate HL, a property which, as we saw, cdn's can also instantiate. To wit, when we say that cdn's instantiate the property of *horse-lookingness*, what we have in mind is, of course, that *cows* instantiate HL, and HL is surely a "contingent"property of cows. That is, it is a property which cows can gain and

[208]I allow that cognitive systems of a different kind, say of Martians, would have a different response modality than humans.

lose quite readily, as is shown by the fact that it is quite an easy feat to distinguish cdn's from horses. If you are in a world (or situation) in which you have confused cdn's with horses, all you have to do is get to a possible world in which you have at your disposal some basic means to make cows look like cows rather than

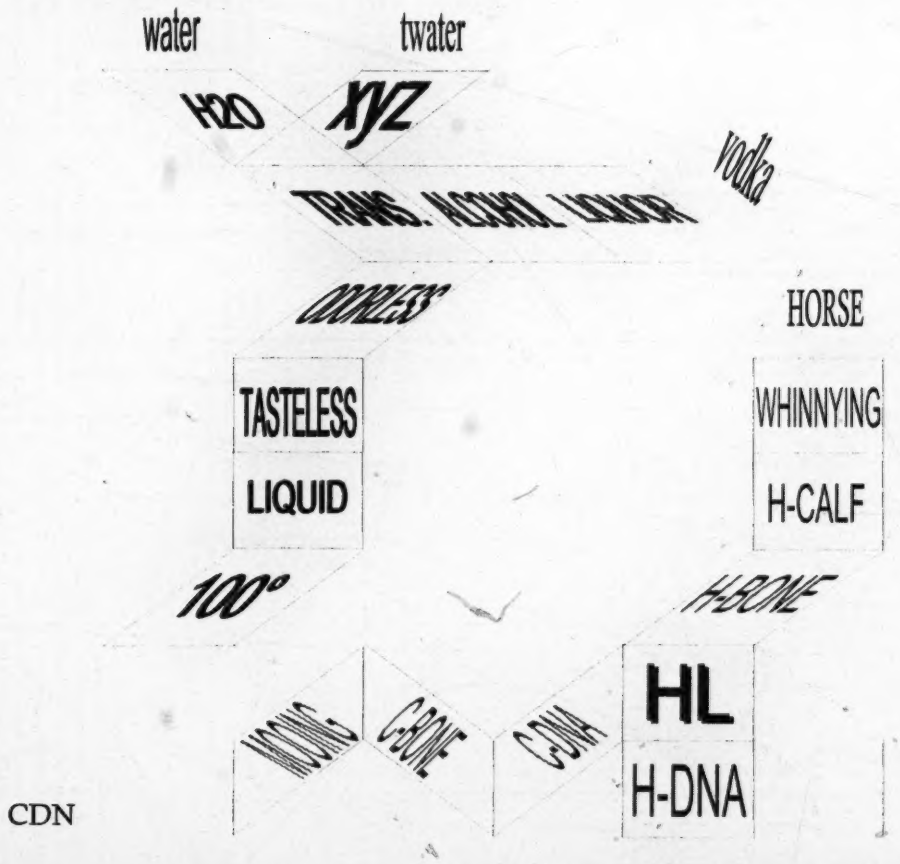


Figure 7.1: A schematic representation of the properties *being water, being twater, being vodka, being horse, and being cdn.* Each rectangle represents one detection property which can cause selective instantiation of the corresponding concepts. As can be seen here, water and twater overlap but for the H2O/XYZ positions, whereas horse and cdn overlap only at the HL position.

horses. As I said, a world in which we have a flashlight will do. This shows, I would argue, that the property of HL, a detection property of horses, is a weakly nomic property of horses, since it seems that even cdn's, which are easy to distinguish from horses, can instantiate it. In contrast, the strong nomic properties of horses would be those where we have to go much "farther" away from the actual world to distinguish horses from other creatures who happen to instantiate them. And this logic generalizes.

For example, suppose that two distinct species of bacteria differ only in a cluster of genes whose detection requires the use of an enzyme that would be available to scientific use only in 500 years from now. Right now we treat these two species as one, although after the discovery of their differences some previously unexplained phenomena, say a certain disease, would receive an explanation. I would claim then that each of these clusters of genes is more of an essential property of this species than that of some other property that easily distinguishes them from other species, perhaps their surface structure.

These considerations seem to show that what makes an ordinary property into an essential property of a thing (or from the perspective of informational theory, a strongly nomic property) is its differential contribution to that thing's selective detection, *relative to a detection system.* Thus, horse-DNA is more of an essential property of horses to humans than to intelligent creatures whose perceptual system is at the micro-level (see chapter Six section 1), since it has a higher "sufficiency value" to *selectively* cause instantiations of HORSE's.[209] Indeed, one would think that strong nomic connections are harder to get by. For non-horses to share the detection property of horse-DNA with horses is a feat which would require much greater technological and theoretical development to discern, than sharing HL with them. And sometimes informational theory acknowledges that the effort might be perhaps not quite worthwhile, to put it crudely.

To sum up then, my claim is that informational semantics is built around the idea of a logical space within which properties can be distinguished to varying degrees along a continuum of gradient of detectability, going from the actual world to the remote areas of distinguishability-for-humans space. The idea is that informational theory carves out sections of these continua when it comes to determine the delineations of its contents. For example, we saw that it does not take too much, in terms of possibilia metrics, to get from the actual world to worlds in which cdn's are distinguishable from horses, which means that making

---

[209]As said in the text, we should distinguish our capacities for selective response from that of other species. Thus, for a species of aliens whose detection capacities are like those of viruses, HL would be more of an essential property of horses than horse-DNA.

the HORSE concept distinct from that of (HORSE v CDN) is quite a trivial feat. On the other hand, to get from ancient times to a world in which water could be distinguished from twater, or for that matter, Carbon 12 from Carbon 14, would be quite, well, *impossible*.[210] Because of this, informational theory leaves the distinctions between the latter kind of properties as moot, for after all, it is the possibilities mostly relevant to the here and now that we should worry about (at least as far as our behavior is concerned), not those at the other end of detectability space. And of course, in between these two extremes one can find all sorts of intermediate cases, such as perhaps the cases of Elm and Beech, Jadeite and Nephrite, etc., in which the kind of expertise or theory required to tell them apart is not "rocket science" as people say (although not trivial either) .

So why is my informational theory not a species of *epistemic optimality* theories, as the theories of Stampe 1975, 1977, and Dretske 1983? On the face of it, it does seem to be just such a theory since it presumes that content is determined by reference to conditions under which agents are able to discriminate between wild and veridical causes of their concepts, namely when they are in optimal epistemic conditions. But as said, these theories are circular since they need to presume that which the agent needs to discriminate, i.e., the content property, in order to specify when the agent is in the optimal epistemic condition. The same is of course true also for other sorts of teleological theories, as that of Millikan, in which the optimal conditions are not necessarily epistemic, just evolutionary optimal (i.e., survival-wise). But my theory is not optimality theory on either version since I have not made any claim about when agents are functioning at their "best". In effect, I can clearly foresee a situation where the "ideal" theory is discovered when agents are functioning not at their "best" but at their "worst", and its discovery was a matter of sheer luck ("Epistemic Luck"). Indeed, the "ideal" theory need not be even *true,* as long as it secures the stipulated covariance relation. And what is more, the agents need not even know that they have *discovered* the "ideal" theory; it is enough that they are able to affect discriminations they could not affect before.

This latter stipulation further shows, I would argue, why the theory is not an "optimality" theory or even why it is not an epistemic theory in the traditional sense. Traditional epistemic theories require that there is someone who *knows* the relevant piece of information, most preferably the agent herself, and who can specify it when required. In the limiting case, it is the theoretician who produces the semantic theory who has to know, to the very least, in what the optimality conditions consist. Hence the inevitable circularity of such theories. But no such

[210] We might just try to imagine how different the history of civilization had to be so that molecular chemistry, or quantum physics, would have been developed already in Plato's time.

thing is required in my theory. First, the theory does not stipulate optimality conditions. Agents can stand in the right, i.e., discriminative, covariance relations to property instantiations even when they are not functioning at their epistemic "best". The idea that false theories can be used just as well as true ones to establish reference shows this I think. Second, our concept of X is a function of our *possible* discriminative abilities. Specifying here and now our possible discriminative abilities, as versions of traditional epistemic theories require, is not only *impossible*, but would lead to a contradiction. For in such a case these possible capacities will become actual. We just have to accept that some things we cannot know "in advance" even with respect to our own concepts. That was Descartes' mistake.

Bibliography

Ackerman, D.F. (1986), "Essential Properties and Philosophical Analysis", in P. French T. Uehling, and H. Wettstein, (eds.), *Midwest Studies in philosophy*, vol. XI (Minneapolis: University of Minnesota Press), pp. 315-29.

Ackerman, D.F. (1990), "The Second Paradox of Analysis", in *Philosophical Perspectives,* vol. 4, pp. 535-543.

Almog, J. (1984), "Semantical Anthropology", in P. French, T. Uehling, and H. Wettstein, (eds.), *Midwest Studies in philosophy,* vol. IX (Minneapolis: University of Minnesota Press), pp. 479-89.

Anscombe, E. (1976), *Intention* (Ithaca, NY: Cornell University Press).

Antony, L. (1989), "Anomalous Monism and the Problem of Explanatory Force", *The Philosophical Review,* vol.98, pp. 153-187.

Antony, L. (1994), "Law and Order in Psychology", *Philosophical Perspectives*, vol. 9, pp. 429-446.

Antony L. and Levine, J. (1991), "The Nomic and the Robust", in Loewer and Rey 1991.

Ayer, J. A. (1946), *Language Truth and Logic* (London: Golancz, 2nd ed.).

Baker, L. R. (1995), *Explaining Attitudes: A Practical Approach to Mind*, (Cambridge, NY: Cambridge Studies in Philosophy).

Barwise, J. and Perry, J. (1981), "Semantic Innocence and Uncompromising Situations", in P. French, T. Uehling, and H. Wettstein, (eds.), *Midwest Studies in the philosophy of Language,* vol. VI (Minneapolis: University of Minnesota Press).

Barwise, J. and Perry, J. (1983), *Situations and attitudes* (Cambridge: MIT).

Bealer, G. (1982), *Quality and Concept* (Oxford: Oxford University Press).

Bealer, G. (1993a), "Universals", *The Journal of Philosophy*, vol. 90, pp. 5-32.

Bealer, G. (1993b), "A Solution to Frege's Puzzle", *Philosophical Perspectives*, 7, pp.17-59.

Bealer, G. (1994), "Mental Properties", *Journal of Philosophy*, vol. 91, pp. 185-208.

Bealer, G. (1998), "Propositions", *Mind* vol, 107, issue 425, pp. 1-32.

Block, N. (1980), "Introduction: What is Functionalism", in *Readings in Philosophy of Psychology*, N. Block, (ed.), Vol. 1 (Cambridge, Mass: Harvard University Press).

Block, N. (1986), "Advertisement for a Semantics for Psychology", in *Midwest Studies in Philosophy*, Vol. X: *Studies in the Philosophy of Mind*, P. French, T. Uehling, and H. Wettstein, (eds.), (University of Minnesota Press, Minneapolis), pp. 615-78.

Boghossian, P. A. (1991), "Naturalizing Content", in Loewer and Rey 1991, pp. 65-86.

Boghossian, P. A. (1993), "Does Inferential Role Semantics Rest Upon a Mistake?", *and Language* , vol. 8, pp. 27-40.

Brentano, F. (1874), "The Distinction Between Mental and Physical Phenomena", in *Psychology from the Empirical Standpoint*, trans. by D.B. Terrell, in Chisholm 1960.

Burge, T. (1979), "Individualism and the Mental",in P. French, T. Uehling, H. Wettstein, (eds.), in *Midwest Studies in Philosophy*, vol. IV, *Studies in Epistemology*, (Minneapolis: University of Minnesota Press).

Burge, T. (1982), "Other Bodies", A. Woodfield, (ed.), *Thought and Object*, (Clarendon Press) pp. 97-120.

Burge, T. (1986), "Individualism and Psychology", *Philosophical Review*, vol. 95, pp. 3-45.

Burge, T. (1988), " Individualism and Self-Knowledge", *Journal of Philosophy*, vol. 85, pp. 649-63.

Burge, T. (1989), " Individuation and Causation in Psychology", in *Pacific Philosophical Quarterly*, vol. 70, pp. 303-22.

Burge, T. (1992), "Philosophy of Language and Mind: 1950-1990", Philosophical Review 100, pp. 3-51.

Burge, T. (1993), "Mind- Body Causation and Explanatory Practice", in Heil and Mele 1993, pp. 97-120.

Butler, T. (1996), "Content, Causal Powers, and Context", *Philosophy of Science*, vol. 63, issue 1, pp. 105-114.

Carlson N. R. (1999), *Foundations of Physiological Psychology* (Allyn and Bacon).

Carnap, R. (1945), "On Inductive Logic", *Philosophy of Science*, vol.12,no. 2, pp. 72-97.

Carnap, R. (1947), *Meaning and Necessity* (Chicago: University of Chicago Press).

Chisholm, R. (1957), *Perceiving: a Philosophical Study* (NY: Cornell University Press).

Chisholm, R. (1960), *Realism and the Background of Phenomenology* (Glencoe Ill).

Chisholm, R. (1967), "Intentionality", in *Edwards*, vol. 4, pp. 201-204.

Chomsky, N. (1959), "A Review of B.F. Skinner's Verbal Behavior", in *Language,* vol. 35, no. 1, pp. 26-58.

Coffa, J. A. (1973), *The Foundations of Inductive Explanation*, Doctoral Dissertation, (University of Pittsburgh).

Cummins, R. (1983), *The Nature of Psychological Explanation* (Cambridge, Mass., MIT Press).

Cummins, R. (1996), *Representations, Targets and Attitudes* (Cambridge, Mass., MIT Press).

Cummins, R. (2000), "Reply to Millikan", *Philosophy and Phenomenological Research* vol. 60, no. 1, pp. 113-127.

Church, A. (1943), "Carnap's Introduction to Semantics", in *Philosophical Review*, vol. 52, pp. 298-305.

Church, A. (1954), "Intensional Isomorphism and Identity of Belief", *Philosophical Studies*, V, pp. 56-73.

Church, S. (1956), *Introduction to Mathematical Philosophy* (Princeton: Princeton University Press).

Churchland, M. (1979), *Scientific Realism and the Plasticity of Mind* (Cambridge University Press).

Churchland, M. (1981), "Eliminative Materialism and the Propositional Attitudes", in *Journal of Philosophy*, vol. 78, no. 2, pp. 67-90.

Churchland, M. and Churchland P.M. (1983), "Stalking the Wild Epistemic Engine", *Nous*, vol. 17,pp. 5-20.

Davidson, D. (1963), "Actions, Reasons, and Causes", in Davidson 1980, pp. 3-19.

Davidson, D. (1969a), "On Saying That", in Davidson 1984, pp. 93-108.

Davidson, D. (1969b), "The Individuation of Events", in Davidson 1980, pp. 181-188.

Davidson, D. (1970), "Mental Events", in *Essays on Actions and Events*, pp. 207-28.

Davidson, D. (1973), "The Material Mind" in Davidson 1980, pp. 245-259.

Davidson, D. (1974), "On the Very Idea of a Conceptual Scheme", in Davidson 1984, pp. 183-98.

Davidson, D. (1980), *Essays on Actions and Events* (Oxford: Clarendon Press).

Davidson, D. (1984), *Inquiries Into Truth and Interpretation* (Oxford: Clarendon Press).

Davidson, D. (1984a), "Radical Interpretation", in *Inquiries Into Truth and Interpretation*, pp. 125-39.

Davidson, D. (1984b), "Truth and Meaning", in *Inquiries Into Truth and Interpretation*, pp. 17-36.

Davidson, D. (1984c), "Belief and the Basis of Meaning", in *Inquiries Into Truth and Interpretation* (Oxford: Clarendon Press), pp. 141-154.

Davidson, D. (1985), "Replies to Essays X-XII", in *Essays on Davidson: Actions and Events*, B. Vermazen, and E. M. Hintikka,(ed.), (Oxford: Clarendon Press).

Davidson, D. (1986a), "A Coherence Theory of Truth", in E. LePore, (ed.), *Truth and Interpretation* (Oxford: Blackwell), pp. 433-46.

Davidson, D. (1986b), "A Nice Derangement of Epitaphs", in E. LePore, (ed.), *Truth and Interpretation* , (Oxford: Blackwell), pp. 433-46.

Davidson, D. (1987), "Knowing One's Own Mind", *Proceedings of the American Philosophical Association*, vol. 60, pp. 441-58.

Davidson, D. (1990), "The Structure and Content of Truth", *Journal of Philosophy*, vol. 87, pp. 279-328.

Davidson, D. (1993), "Thinking Causes", in J. Heil, and A. Mele, (eds.), *Mental Causation*, (Oxford: Clarendon Press), pp. 3-17.

Davidson, D. and Hintikka, J. (1969), *Words and Objections* (Dordrecht: D. Reidel).

Davies, M. (1991), "Individualism and Perceptual Content", *Mind*, vol. C, issue 4, pp. 461-484.

Dennett, D. (1991a), "Real Patterns", *Journal of Philosophy*, vol. 88, no. 1, pp. 27-51.

Dennett, D. (1991b), "Mother Nature Versus the Walking Encyclopedia: Western Drama", in *Philosophy and Connectionist Theory*, W. Ramsey, S. Stich, and D. Rumelhart, (eds.), (Lawrence Erlbaum Associates).

Devitt, M. (1990), "A Narrow Representational Theory of the Mind", in Lycan 1990.

Devitt, M. (1996), *Coming to Our Senses: A Naturalistic Program for Semantic Localism.* (Cambridge: Cambridge University Press).

Donnellan, K. (1966), "Reference and Definite Descriptions", *Philosophical Review*, vol. 75, pp. 281-304.

Donnellan, K. (1970), "Proper Names and Identifying Descriptions", *Synthese* , vol. 21, pp. 335-358.

Donnellan, K. (1978), "Speaker Reference, Descriptions and Anaphora", in P. Cole, (ed.), *Syntax and Semantics*, vol. 9, (New-York: Academic Press), pp. 47-68.

Dray, H. W. (1957), *Laws and Explanation in History* (Oxford, UK. Oxford University Press).

Dray, H. W. (1963), "Historical Explanation of Action Reconsidered",in *The Philosophy of History*, P. Gardiner, (Ed.), (Oxford, UK, Oxford University Press).

Dretske, F. (1969), *Seeing and Knowing* (University of Chicago Press).

Dretske, F. (1977), "Laws of nature", in *Philosophy of Science*, vol. 44, pp. 248-68.

Dretske, F. (1979), "Simple Seeing", in *Body, Mind and Method,* D.F. Gustafson, (ed.), (Dordrecht: Reidel), pp.1-16.

Dretske, F. (1981), *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press).

Dretske, F. (1983a), "The Epistemology of Belief", in *Synthese*, vol. 55, pp. 3-19.

Dretske, F. (1986), "Misrepresentation", in R. F. Bogdan, (ed.), *Belief: Form, Content, and Function*, (Clarendon Press), pp.17-36.

Dretske, F. (1995), *Naturalizing the Mind* (Cambridge, MA: MIT Press).

Dummett, M. (1973), *Frege, Philosophy of Language* (London: Duckworth).

Dummett, M. (1975), "What is a Theory of Meaning? (I)", in Guttenplan, (ed.), *Mind and Language*, (Oxford: Oxford University Press), pp. 97-138.

Dummett, M. (1976), "What is a Theory of Meaning (II)", in G. Evans, and J. McDowell.

Dummett, M. (1978),"Truth", in *Truth and Other Enigmas* (Cambridge: Harvard University Press).

Edgington, D. (1985), "The Paradox of Knowability", *Mind*, New Series, vol. 94, issue 376, pp. 557-568.

Edwards, P. (1967), (ed.) *The Encyclopedia of Philosophy* (NY: Macmillan).

Egan, F. (1991), "Must Psychology Be Individualistic", *Philosophical Review,* vol. 100, pp. 179-203.

Egan, F. (1995), "Computation and Content", *The Philosophical Review*, vol. 104, no. 2., pp. 181-203.

Evans, G.(1973), "The Causal Theory of Names", *Aristotelian Society: Supplementary* vol. 47, pp. 187-208.

Evans, G.(1982), *The Varieties of Reference* (Oxford: Clarendon Press).

Evans, G. and McDowell, J. (1976), (eds.), *Truth and Meaning: Essays in Semantics* (Oxford: Oxford University Press).

Feigl, H. (1958), "The 'Mental' and the 'Physical'", in *Minnesota Studies in the Philosophy of Science*, vol. II, (Minneapolis: University of Minnesota Press), pp. 370-497.

Field, H. (1972), "Tarski's Theory of Truth", *Journal of Philosophy*, vol. 69, pp. 347-75.

Field, H. (1977), "Logic Meaning and Conceptual Role", *Journal of Philosophy*, vol.74, pp. 379-409.

Field, H. (1978), "Mental Representation", *Erkenntnis*, vol. 13, pp. 9-61.

Fields, B. N. (1990), (ed). *Virology*. (New-York: Raven Press).

Fodor, J. (1965), "Explanation in Psychology", in *Philosophy in America*, M. Black, (ed.), (Muirhead Library of Philosophy).

Fodor, J. (1968a), "The Appeal to Tacit Knowledge in Psychological Explanation", *Journal of Philosophy*, vol. 65, pp. 627-640.

Fodor, J. (1968b), *Psychological Explanation* (NY: Random House).

Fodor, J.(1974), "Special Sciences", *Synthese*, vol. 28, pp. 97-115.

Fodor, J.(1975), *The Language of Thought* (NY: Crowell, 1975).

Fodor, J.(1978), "Propositional Attitudes", *The Monist*, vol. 61, no. 4, pp. 501-523.

Fodor, J.(1980), "Methodological Solipsism Considered as a Research Strategy in Cognitive Science", *Behavioral and Brain Sciences*, vol. 3, pp. 63-109.

Fodor, J.(1981), *Representations* (Cambridge, MA: MIT Press).

Fodor, J. (1982), "Cognitive Science and the Twin-Earth Problem", *Notre Dame Journal of Formal Logic*, vol. 23, pp. 98-118.

Fodor, J. (1983), *The Modularity of Mind: An Essay on Faculty Psychology* (MIT press, Bradford books).

Fodor, J. (1984a), "Semantics, Wisconsin Style", *Synthese*, vol. 59, pp. 231-250.

Fodor, J. (1984b), "Observation Reconsidered", *Philosophy of Science*, vol. 51, pp. 23-43.

Fodor, J. (1986a), "Individualism and Supervenience" in *Proceedings of the Aristotelian Society Supplementary*, vol. 60, pp. 235-262.

Fodor, J. (1986b), "Why Paramecia don't have Mental Representations", in P. French, T. Uehling, and H. Wettstein, (eds.), *Midwest Studies in philosophy*, vol. X, (Minneapolis: University of Minnesota Press).

Fodor, J. (1987), *Psychosemantics: The problem of Meaning in the Philosophy of Mind* (Cambridge MA: Bradford).

Fodor, J. (1989b), "Making Mind Matter More", in *Philosophical Topics*, vol. 67, no. 1 , pp. 59-79.

Fodor, J. A. (1990a), "Information and Representation", in P. Hanson, (ed.), *Information, Language, and Cognition* , (Vancouver: University of British Columbia Press), pp. 175-190.

Fodor, J. (1990b), *A Theory of Content and Other Essays* (Cambridge, MA: MIT Press).

Fodor, J. (1991), "A Modal Argument for Narrow Content", *Journal of Philosophy*, vol. 88, pp. 5-26.

Fodor, J. (1994), *The Elm and the Expert* (Cambridge, Mass. MIT Press).

Fodor, J. (1997), "Special Sciences: Still Autonomous After All These Years", in *Philosophical Perspectives* 11, *Mind ,Causation, and the World*, J. E. Tomberlin, (ed.), pp. 149-163.

Fodor, J. (1998), *Concepts: Where cognitive science went wrong* (Oxford, Clarendon Press).

Fodor, J. and LePore E. (1992), *Holism: A Shopper's Guide* (Oxford, Blackwell).

Fodor, J. and McLaughlin B. (1990), "Connectionism and the Problem of Systematicity: why Smolensky's Solution doesn't work, *Cognition*, vol. 35, pp. 183-204.

Fodor, J. and Pylyshyn Z. (1988), "Connectionism and Cognitive Architecture: a Critical Analysis", *Cognition*, vol. 28, pp. 3-71.

Frege, G. (1884), *Die Grundlagen der Arithmetic* (*The Basic Laws of Arithmatics*), in M. Furth (trans. and ed.), (Berkeley and Los Angeles: University of California Press, 1964).

Frege, G. (1892a), "On Concept and Object", in P. Geach, and M. Black, (eds.), *Translations from the Philosophical Writings of G. Frege*, 2nd (ed.) (Oxford: Blackwell, 1952).

Frege, G. (1892b), "On Sense and Reference", in P. Geach, and M. Black, (eds.), *Translations from the Philosophical Writings of G. Frege*, 2nd (ed.) (Oxford: Blackwell, 1952), pp. 56-78.

Frege, G. (1918), "The Thought: A Logical Inquiry", in P. Strawson, (ed.), *Philosophical Logic*, Tr. by A. M. and M. Quinton , (Oxford: Oxford University Press, 1967).

Frege, G. (1919-23), *Logical Investigations*, Tr. by P. Geach, (Oxford: Blackwell, 1977).

Frege, G. (1952), *Translations from the Philosophical Writings of G. Frege*, 2nd (ed.) P. Geach, and M. Black, (eds.), (Oxford: Blackwell, 1952).

Frege, G. (1980), *Gottlob Frege: Philosophical and Mathematical Correspondence*, G. Gabriel, H. Hermes, F. Kambartel, C. Theil, and A. Veraart (eds.), Tr. by H. Kaal, (Chicago: The University Press of Chicago Press).

Gödel, K. (1944), "Russell's Mathematical Logic", in P.A. Schillp, (ed.), *The Philosophy of Bertrand Russell* (Evanston and Chicago: Northwestern University Press).

Goldman, (1976), "Discrimination and Perceptual Knowledge",in *The Journal of Philosophy*, vol. 73, issue 20, pp. 771-791.

Goldman, (1988), "Strong and Weak Justification", in *Philosophical Perspectives*, J.E. Tomberlin, (ed.), pp. 51-69.

Goodman, N. (1955), *Fact, Fiction, and Forcast* (Cambridge, Mass).

Goodman, N. (1978), *Ways of World Making* (Brighton: Harvester).

Grice, H. (1957), "Meaning", in *Philosophical Review*, vol. 66, pp. 377-388.

Grimm, R. and D. Merrill (1988), (eds.), *Content of thought*, (Tucson, Arizona : the University of Arizona Press).

Harman, G. (1973), *Thought* (Princeton, N.J.: Princeton University Press).

Harman, G. (1982), "Conceptual Role Semantics", *Notre Dame Journal of Formal Logic*, vol. 28, pp. 242-56.

Heathcote, A. and D. M. Armstrong, (1991), "Causes and Laws", in *Nous*, vol. 25, no. 1, pp. 63-73.

Heil, J. and Mele, A. (1993), *Mental Causation* (Oxford: Claredon Press).

Hempel, C. (1965), *Aspects of Scientific Explanation* (New York: The Free Press).

Hempel, C. (1970), "Reduction: Ontological and Linguistic Facets", in S. Morgenbesser, P. Suppes and M. White, (eds.), *Essays in the Honor of Earnest Nagel* (NY: St. Martin's Press), pp. 179-199.

Hempel, C. and Oppenheim (1948), "Studies in the Logic of Explanation", in *Philosophy of Science*, vol. 115, pp. 135-146.

Horgan, T. (1978), "Supervenient Bridge-Laws", *Philosophy of Science*, vol. 45, pp. 227-249.

Horgan, T. (1989), " Mental Causation", *Philosophical Perspectives*, vol. 3, pp. 47-76.

Horgan, T. (1993), " From Supervenience to Superdupervenience: Meeting the Demands of a Material World", *Mind*, vol. 102, issue 408, pp. 555-586.

Jackson, F. (1994), "Finding The Mind in The Natural World", in R. Casati, E. Smoth, and G. White, (eds.), *Philosophy and the Cognitive Sciences* (Vienna: Holder-Pichler-Tempski).

Jackson, F. (1996), "The Primary Quality View of Color", in *Philosophical Perspectives* 10, Metaphysics, vol.30, pp. 199-219.

Jackson, F. and Pettit, P. (1988), "Functionalism and Broad Content", *Mind*, vol. 97, issue 387, pp. 381-400.

Jackson, F. and Pettit, P. (1990), "Causation in the Philosophy of Mind" *Philosophy and Phenomenological Research*, vol. 1, pp. 195-214.

Johnston, M. (1993), "Objectivity Refigured: Pragmatism Without Verificationism" in J. Haldane, C. Wright, (eds.), *Reality, Representation and Projection*, (New-York: Oxford University Press).

Kaplan, D. (1978), "Dthat", in P. Cole, (ed.), *Syntax and Semantics*, vol. 9 (New-York: Academic Press), pp. 221-253.

Kaplan, D. (1980), "Demonstratives", in J. Almog, et al, (eds.), *Themes from Kaplan,* (New-York: Oxford University Press).

Katz, J. J. (1972), *Semantic Theory* (New-York: Harper and Row).

Kim, J. (1966), "On the Psycho-Physical Identity Theory", in *American Philosophical Quarterly*, vol. 3, pp. 227-235.

Kim, J. (1973), "Causation, Nomic Subsumption and the Concept of Event", *Journal of Philosophy*, (reprinted in Kim 1993a), vol. 70, pp. 217-236.

Kim, J. (1976), "Events as Property Exemplification", in *Action Theory*, M. Brand and D. Walton, (Eds.), (Dordrecht, Holland: D. Reidel), Reprinted in Kim 1993, pp. 159-177.

Kim, J. (1982), "Supervenience and Microphysics", *Pacific Philosophical Quarterly*, vol. 63, pp. 29-43.

Kim, J. (1983), "Supervenience and Supervenient Causation" in *Southern Journal of Philosophy*, supp., vol. 22, pp. 45-56.

Kim, J. (1984a), "Concepts of Supervenience", *in Philosophy and Phenomenological Research*, vol. 45, pp. 153-177, Reprinted in Kim 1993.

Kim, J. (1984b), "Epiphenomenal and Supervenient Causation", in P. French, T. Uehling, and H. Wettstein, (eds.), *Midwest Studies in philosophy*, vol. IX, (Minneapolis: University of Minnesota Press), pp. 257-270.

Kim, J. (1989), "Mechanism, Purpose, and Explanatory Exclusion", *Philosophical Perspective*, vol. 3, pp. 77-108.

Kim, J. (1992), "Multiple Realization and the Metaphysics of Reduction", *Philosophy and Phenomenological Research*, vol. LII, pp. 1-26.

Kim, J. (1993a), *Supervenience and Mind*, (Cambridge: Cambridge University Press).

Kim, J. (1993b), "The Non-Reductivist's Trouble with Mental Causation", in Heil and Mele 1993, pp. 189-210.

Kim, J. (1994), "Supervenience", in *A Companion to the Philosophy of Mind*, S. Guttenplan, (ed.), (Cambridge, MA: Blackwell), pp. 575-583.

Kim, J. (1998), *Philosophy of Mind* (Oxford: Westview Press).

Kim, K. M. (1994), *Explaining Scientific Consensus: The Case of Mendelian Genetics* (NY, Guilford Press).

Kohler, W. (1929), *Gestalt Psychology* (NY, Liveright).

Kohler, W. (1942), *Dynamics in Psychology* (London, Faber&Faber).

Kripke, S. (1972), *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980).

Kripke, S. (1979), "A Puzzle about Belief", in A. Margalit, (ed.), *Meaning and Use* (Dordrect: D. Reidel, 1979), pp. 239-283.

Kripke, S. (1982), *Wittgenstein on Rules and Private Language* (Harvard University Press).

Kuhn, T. (1962), *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press).

LePore, E and Loewer, B. (1987), "Mind Matters", in *Journal of Philosophy*, vol. 84, pp. 630-641.

LePore, E. and Loewer, B. (1989), "More on Making Mind Matter", *Philosophical topics*, vol. 17, pp. 175 -192.

Lewis, D. (1970), "How to Define Theoretical Terms", in *Journal of Philosophy*, vol. 67, pp. 427- 446, Reprinted in Lewis 1983b.

Lewis, D. (1972), "Psycho-Physical and Theoretical Identifications", *Australasian Journal of Philosophy*, vol. 50, pp. 249-258.

Lewis, D. (1973), *Counterfactuals* (Cambridge, Mass: Harvard University Press).

Lewis, D. (1983a), "Extrinsic Properties", *Philosophical Studies*, vol. 44, pp.197-200.

Lewis, D. (1983b), *Philosophical Papers*, vol. I (Oxford: Oxford University Press).

Lewis, D. (1986a), *On the Plurality of Worlds* (Oxford: Blackwell).

Lewis, D. (1986b), *Philosophical Papers*, vol. II (Oxford: Oxford University Press).

Linsky, L. (1963), "Reference and Referents", in C. Caton, (ed.), *Philosophy and Ordinary Language* (Urbana, Ill: University of Illinois Press, 1963).

Linsky, L. (1977), *Names and Descriptions* (Chicago: Chicago University Press).

Loar, B. (1981), *Mind and Meaning* (Cambridge: CUP, 1981).

Loar, B. (1982), "Conceptual Role and Truth-Conditions", in *Notre Dame Journal of Formal Logic*, vol. 28, pp. 272-283.

Loar, B. (1985), "Social Content and Psychological Content", in Grimm and Merrill 1988, pp. 99-110.

Loewer, B. and Rey, G. (1991), *Meaning in Mind: Fodor and his Critics* (Oxford, Blackwell).

Lycan, W.G. (1981),"Psychological Laws", in *Philosophical Topics*, vol. 12, no.1, pp. 9-38.

Lycan, W.G. (1990), *Mind and Cognition* (Cambridge MA: Basil Blackwell).

Marr, D. (1982) *Vision* (San Francisco: Freeman and Company).

Mates, B. (1951), "Synonymity", in University of California Publications in Philosophy, vol. 25, pp. 201-26. Reprinted in *Semantics and the Philosophy of Language*, L. Linsky ,(ed.) , University of Chicago Press, 1952.

McDowell, J. (1977),"On the Sense and Reference of a Proper Name", *Mind*, vol. 86, pp. 159-185.

McDowell, J. (1986), "Singular Thought and the Extent of Inner Space", in Pettit and McDowell, pp. 137-168.

McDowell, J. (1994), *Mind and World* (Cambridge, MA: Harvard University Press).

McGinn, C. (1982), "The Structure of Content", in *Thought and Object*, A. Woodfield,(ed.), (Oxford University Press), pp. 207-58.

McGinn, C. (1989), "Can we Solve the Mind-Body Problem?", in *Mind*, vol. 98 pp. 349-366.

McGinn, C. (1989a), *Mental Content* (Oxford: Blackwell).

McGinn, C. (1991), "Conceptual Causation: Some Elementary Reflections", *Mind*, vol. 100, pp. 573-586.

McLaughlin, B. P. (1989), "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical", in *Philosophical Perspectives*, vol. 3, pp. 109-135.

McLendon, H. (1955), "Uses of Similarity of Structure in Contemporary Philosophy", *Mind* New Series, vol. 64, issue 253, pp. 79-95.

Melden, A. I. (1961), *Free Action* (Boston, Mass., Routledge and P. Keagan).

Mill, J. (1843), "Of Names", in J. Robson, (ed.), *A System of Logic* (Toronto: University of Toronto Press), pp. 24-45.

Millikan, R. (1984), *Language, Thought and other Biological Categories* (Cambridge, Mass: MIT Press).

Millikan, R. (1986), "Thought Without Laws: Cognitive Science with Content", in *Philosophical Review*, vol. 95, pp. 47-80.

Millikan, R. (1989), "Biosemantics", in *Journal of Philosophy*, (Reprinted in Stich, and Warfield, 1994), vol. 86, pp. 281-297.

Moore, G.E. (1952), "Analysis: A reply to My Critics", in W. Alston, and G. Nakhnikian, (eds.), *Readings in Twentieth Century Philosophy* (London, Free Press), pp. 279-285.

Nagel, E. (1950), *The Structure of Science* (London: Routledge, and P. Kegan, 1961).

Neale, S. (1990), "Descriptive Pronouns and Donkey Anaphora", *The Journal of Philosophy*, vol. 87, pp. 113-150.

Neale, S. (1995), "The Philosophical Significance of Gödel's Slingshot", *Mind*, vol. 104, pp. 761-825.

Noonan, H.W. (1991), "Object-Dependent Thoughts and Philosophical Redundancy", *Analysis* 51, pp, 1-16.

Oppenheim, P. and Putnam, H. (1958), "Unity of Science as a Working Hypothesis", in H. Feigl, M. Scriven, & G. Maxwell, (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis: University of Minnesota Press).

Papineau, D. (1993), *Philosophical Naturalism* (Oxford: Blackwell).

Peacocke, C. (1983), *Sense and Content: Experience, Thought, and their Relations* (Oxford, Clarendon Press).

Peacocke, C. (1988), "The Limits of Intelligibility: A Post Verificationist Proposal" *Philosophical Review*, vol. 97, no. 4, pp. 463-496.

Peacocke, C. (1989), "What are Concepts?", in *Midwest Studies in Philosophy*, vol. 14, pp. 1-28

Peirce, C. S. (1931), *Collected Papers* (Cambridge, Mass, Harvard University Press).

Perry, J. (1977), "Frege on Demonstratives", in *The Philosophical Review*, vol. 86, pp. 474-97.

Perry, J. (1979), "The Problem of the Essential Indexical", *Nous*, vol. 13, pp. 3-21.

Perry, J. (1994) "Intentionality (2)" in Guttenplan, Samuel (ed.) *A Companion to the Philosophy of Mind* (Oxford, Blackwell), pp. 386-395.

Perry, J. (1997) "Broadening the Mind: Review of Jerry Fodor's the Elm and the Expert" in *Philosophy and Phenomenological Research*, pp. 1-8.

Perry, J. and Israel, D. (1991), "Fodor and Psychological Explanations", in Loewer, and Rey, pp. 165-180.

Pettit, P. (1991), "Realism and Response-Dependence", *Mind*, New Series, vol. 100, issue 4 pp. 587-626.

Pettit, P. and J. McDowell (1986), *Subject, Thought, and Context* (Oxford: Clarendon Press).

Place, U. (1956), "Is Consciousness a Brain Process*?*", in *British Journal of Psychology*, vol. XLVII, pp. 44-50.

Putnam, H. (1960), "Mind and Machines", in S. Hook (ed.), *Dimensions of Mind*, (NY, New-York University Press) pp. 138-164.

Putnam, H. (1965), "Brains and Behavior", in *Analytical Philosophy*, vol. 2, R. Butler, (ed.) (Oxford: Blackwell).

Putnam, H. (1967), "Psychological Predicates", in *Arts, Philosophy and Religion* (Pittsburgh: University of Pittsburgh Press). Later reprinted under the title "The Nature of Mental States".

Putnam, H. (1973) "Explanation and Reference", in G. Pears and P. Maynard (eds.) *Conceptual Change* (Reidel).

Putnam, H. (1975), "The Meaning of 'Meaning'", in K. Gunderson, (ed.), *Language, Mind & Knowledge* (Minneapolis, MW: University of Minnesota Press), pp. 131-193.

Putnam, H. (1981), *Reason, Truth and History* (Cambridge: Cambridge University Press).

Putnam, H. (1992), *Representation and Reality* (Cambridge, Mass., MIT Press).

Putnam, H. (1990), *Realism With a Human Face* (Cambridge, Mass., Harvard University Press).

Putnam, H. (1994), *Words and Life* (Cambridge, Mass: Harvard University Press).

Pylyshyn, Z. (1980), "Computation and Cognition: Issues in the Foundations of Cognitive Science", Behavioral and Brain Sciences, vol. 3, pp. 111-132.

Quine, W. (1953), "Two Dogmas of Empiricism", in *From a Logical Point of View* (New York: Harper and Row), pp. 20-46.

Quine, W. (1960), *Word and Object* (Cambridge, MA: MIT Press).

Quine, W. (1969), *Ontological Relativity and Other Essays* (New York: Columbia University Press).

Quine, W. (1974), *The Roots of Reference* (Ill.: Open Court).

Quine, W. (1990), *Pursuit of Truth* (Cambridge, MA: Harvard University Press).

Ramsey, F. (1929), "Theories", in D. Mellor, (ed.), *Foundations* (London: Rutledge and Kegan Paul).

Richard, M. (1990), *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them* (New-York: Cambridge University Press).

Rorty, R. (1980), *Philosophy and the Mirror of nature* (Oxford: Basil Blackwell).

Russell, B.(1903), *Principles of Mathematics* (New York: W. Norton).

Russell, B. (1905), "On Denoting", in R. Marsh, (ed.), *B. Russell, Logic and Knowledge* (London: Allen and Unwin 1956), pp. 41-56.

Russell, B. (1918), "The Philosophy of Logical Atomism", in Russell 1956, pp. 175-281.

Russell, B. (1919), "Descriptions", in *Introduction to Mathematical Philosophy* (London: Allen and Unwin, 1919), pp. 167-180.

Russell, B. (1929), "Knowledge by Acquaintance and Knowledge by Description", in *Mysticism and Logic* (NY: W. Norton).

Russell, B. (1948), *Human Knowledge* (NY: Simon & Schuster).

Russell, B. (1956), *Logic and Knowledge* (Worchester: Unwin Hyman Ltd.).

Ryle, G. (1949), *The Concept of Mind* (New-York: Barnes and Noble).

Salmon, N. (1981), *Reference and Essence* (Princeton, NJ: Princeton University Press).

Salmon, N. (1986/91), *Frege's Puzzle Atascadero* (California: Ridgeview).

Salmon, N. (1990), "Reference and Information Content: Names and Descriptions", in D. Gabbay, and F. Guenthner, (eds.), *Handbook of Philosophical Logic IV: Topics in the Philosophy of Language* (Dordrect: D. Reidel,1989), pp. 409-461.

Schiffer, S. (1981), "Truth and the Theory of Content", in H. Parret, and J. Bouvaresse, (eds.), *Meaning and Understanding* (Berlin: Walter de Gruyter).

Schiffer, S. (1982), "Intension-Based Semantics", *Notre Dame Journal of Formal Logic*, vol. 28, pp. 119-56.

Schiffer, S. (1987), *Remnants of Meaning* (Cambridge, Mass: MIT Press).

Schiffer, S. (1991), "Ceteris Paribus Laws", in *Mind*, vol. 100, issue 397, pp. 1-17.

Scheerer, Eckart (1994), "Psychoneural Isomorphism", *Philosophical Psychology*, vol. 7, no. 2, pp. 183-203.

Seager, W. (1992), "Externalism and Token Identity", in *Philosophical Quarterly*, vol. 42 no. 162, pp. 432-448.

Searle, J. (1983), *Intentionality* (Cambridge: Cambridge University Press, 1983).

Searle, J. (1992), *The Rediscovery of the Mind* (Cambridge, Mass., Mit Press).

Segal, G. (1989) "The Return of the Individual", in *Mind* vol. XCVIII no. 389, pp. 39-57.

Sellars, W. (1956), "Empiricism and the Philosophy of Mind", in H. Feigl, and M. Scriven, (eds.), *Minnesota Studies in the Philosophy of Science,* vol. 1 (Minneapolis: University of Minnesota Press), pp. 253-329.

Shaffer, J. (1963), "Mental Events and the Brain", in *Journal of Philosophy,* vol. 60, pp. 160-165.

Shoemaker, S. (1979), "Identity, Properties, and Causality", reprinted in Shoemaker 1984.

Shoemaker, S. (1980), "Causality and Properties", in P. Van Inwagen, (ed.), *Time and Cause* (Dordrecht: Reidel, 1980).

Shoemaker, S. (1984), *Identity, Cause and Mind* (New-York: Cambridge University Press).

Skinner, B. F. (1953), *Science and Human Behavior* (New-York: MacMillan).

Sklar, L. (1967), "Types of Inter-Theoretic Reduction", *British Journal for the Philosophy of Science*, pp. 109-124.

Smart, J. (1959), "Sensations and Brain Processes", *Philosophical Review,* vol. 68, pp. 141-156.

Sosa, E. (1984), "Mind-Body Interaction and Supervenient Causation", in P. French, T. Uehling, and H. Wettstein, (eds.), *Midwest Studies in philosophy*, vol. IX (Minneapolis: University of Minnesota Press), pp. 270-281.

Stalnaker, R. (1976), "Propositions", in *Issues in the Philosophy of Language,* (eds.) Alfred F. MacKay, and D. D. Merrill, (London and New Haven: Yale University Press), pp. 79-91.

Stalnaker, R. (1984), *Inquiry* (Bradford Books, MIT Press).

Stalnaker, R. (1989), "On What's in the Head", in *Philosophical Perspectives,* vol.3, *Philosophy of Mind and Action Theory*, pp. 287-316

Stalnaker, R. (1996), "Varieties of Supervenience", *Supplement: Philosophical Perspectives*, vol. 30, pp 221-241.

Stampe, (1975), "Show and Tell", in B. Freed, A. Marras, P. Maynard (eds.), *Forms of Representation: Proceedings of the 1972 Philosophy Colloquium of the University of Western Ontario* (Amsterdam: North-Holland Publishing Company).

Stampe, (1977), "Toward a Causal Theory of Linguistic Representation", in P. French, T. Uehling, and H. Wettstein, (eds.), *Midwest Studies in Philosophy*, vol. 2, (Minneapolis: University of Minnesota Press).

Stich, S. (1978), "Autonomous Psychology and the Belief-Desire Thesis", *Monist*, vol. 61, pp. 573-591.

Stich, S. (1983), *From Folk Psychology to Cognitive Science: The Case Against Belief* (Cambridge, MA: MIT Press, 1983).

Stich, S. (1992), "What is a Theory of Mental Representation?", *Mind*, vol. 101, issue 402, pp. 243-261.

Stich, S. and Warfield, T. A. (1994), (eds.) *Mental Representation: A Reader* (Cambridge, Blackwell 1994)

Strawson, P. (1950), "On Referring", in *Mind*, vol. 59, issue 235, pp. 320-344.

Strawson, P. (1971), "Propositions, Concepts, and Logical Truth", in *Logico-Linguistic Papers*, (London and NY: Methuen).

Tascheck, W. (1992), "Frege's Puzzle, Sense, and Information Content", in *Mind*, vol. 101 issue 404, pp. 767-791.

Tarski, A. (1944), "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research, 4,* pp. 341-75.

Tarski, A. (1956), "The Concept of Truth In Formalized Languages", in *Logic, Semantics, Mathematics* (Oxford: Oxford University Press).

Tooley, M. (1977), "The Nature of Laws", *Canadian Journal of Philosophy*, vol. 7, pp. 667-98.

Tooley, M. (1987), *Causation: A Realist Approach* (Oxford: Oxford University Press, 1987).

Tye, M. (1992), "Naturalism and the Mental", in *Mind, New Series*, vol.101, issue 403, pp. 421-441.

Vendler, Z. (1972), *Res Cogitans: An Essay in Rational Psychology* (Ithaca: Cornell University Press).

Weitzman, L. (1997), "Frege on the Individuation of Thoughts", *Dialogue*, vol. 36, pp. 563-574.

Wilson, R.A. (1992), "Individualism, Causal Powers, and Explanation", *Philosophical Studies*, vol.68, pp. 103-139.

Wilson, R.A. (1994) "Wide Computationalism" in *Mind* 103(411), pp. 351-372.

Wilson, R.A. (1995), *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind* (NY: Cambridge University Press).

Wilson, R.A. (2001), "Two Views of Realization", in *Philosophical Studies*, vol. 104, pp. 1-31.

Wilson, R.A. (Forthcoming), *The Individual in the Fragile Sciences, vol. 1: Cognition.*

Wittgenstein, L. (1953), *Philosophical Investigations*, trans. by G. E.M. Anscombe (Oxford: Blackwell).